

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A3(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, C12Q 1/68, C07K 16/18, A61K 38/17, C12P
21/00

(21) International Application Number: PCT/IB00/01496

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149.499 18 August 1999 (18.08.1999) US
60/156.503 28 September 1999 (28.09.1999) US(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/156.503 (CIP)
Filed on 18 August 1999 (18.08.1999)(71) Applicant (for all designated States except US): **FRAUN-
HOFER-GESELLSCHAFT ZUR FOERDERUNG
DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE];
Leonrodstrasse 54, D-80636 München (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WIEMANN,
Stefan** [DE/DE]; Grosse Lachstrasse 30a, 69207 Sand-
hausen (DE). **POUSTKA, Annemarie** [DE/DE]; Laden-
burgstrasse 41, 69120 Heidelberg (DE). **WELLEN-
REUTHER, Ruth** [DE/DE]; An der Markscheide 5,
69126 Heidelberg (DE). **BLUM, Helmut** [DE/DE];
Koenigswieser Strasse 94, 81475 Muenchen (DE). **OBER-
MAIER, Brigitte** [DE/DE]; Muehlstrasse 9a, 82547
Eurasberg (DE). **OTTENWAEELDER, Birgit** [DE/DE];
Beinhofstrasse 1a, 81247 Muenchen (DE). **BAHR,
André** [DE/DE]; Raiffaelweg 6, 40724 Hilden (DE).
DUESTERHOEFT, Andreas [DE/DE]; Karlrobert-Kre-
iten-Strasse 14, 40724 Hilden (DE). **KOENIG, Christoph**
[DE/US]; 6233 22nd Avenue N.E., Seattle, WA 98115
(US). **LAUBER, Juergen** [DE/DE]; Unterberg 1F, 42799
Leichlingen (DE). **HEUBNER, Dagmar** [DE/DE];
Grüne Trift 126a, 12557 Berlin (DE). **WAMBUIT,
Rolf** [DE/DE]; Florian-Geyer-Strasse 28, 12489 Berlin
(DE). **KOEHRER, Karl** [DE/DE]; Schlossmannstrasse
4, 40225 Duesseldorf (DE). **BEYER, Andreas** [DE/DE];Helgolandring 106, 45149 Essen (DE). **GASSENHU-
BER, Johann** [DE/DE]; Emanuel Geibel Strasse 8,
65185 Wiesbaden (DE). **GRUBER, Christian** [DE/DE];
Zasinger Strasse 8, 81547 Muenchen (DE). **STRACK,
Norman** [DE/DE]; Linderbergweg 1, 82229 Seefeld
(DE). **MEWES, H.W.** [DE/DE]; Graf Toerring Strasse 9,
82237 Woerthsee (DE). **ANSORGE, Wilhelm** [DE/DE];
Boxberring 107/55, 69126 Heidelberg (DE). **GLASSL,
Sabine** [DE/DE]; Friedberger Weg 2, 64720 Michelstadt
(DE). **RITTMUELLER, Claudia** [DE/DE]; Siedler-
weg 2, 69151 Dilsbergerhof (DE). **REGIERT, Thomas**
[DE/DE]; Raiffeisenstrasse 38, 67227 Frankenthal
(DE). **BLOECKER, Helmut** [DE/DE]; Doeringstrasse
16, 38118 Braunschweig (DE). **BOECHER, Michael**
[DE/DE]; Alter Weg 41a, 38302 Wolfenbuettel (DE).
HORNISCHER, Klaus [DE/DE]; Mozartstrasse 2, 38106
Braunschweig (DE). **NORDSIEK, Gabriele** [DE/DE];
Ohfeld 34, 31188 Holle (DE). **TAMPE, Jens** [DE/DE];
Bergisch-Gladbacher-Strasse 656, 51067 Koeln (DE).(74) Agents: **MERCER, Christopher, Paul** et al.; Carpmals
& Ransford, 43 Bloomsbury Square, London WC1A 2RA
(GB).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
20 June 2002For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.



WO 01/12659 A3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01496

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C12Q1/68 C07K16/18 A61K38/17
C12P21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 09552 A (GENETICS INST) 24 February 2000 (2000-02-24) Page 546, claim 86: SEQ.ID.No.: 77 ---	1-46
X	HILLIER L ET AL: "Human cDNA clone IMAGE:754267" EMBL SEQUENCE DATABASE, 23 July 1997 (1997-07-23), XP002163418 HEIDELBERG DE Accession Nr.: AA478899 abstract --- -/--	1-42



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

20 March 2001

Date of mailing of the international search report

07.06.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HILLIER L ET AL.: "Human cDNA clone IMAGE: 754167" EMBL SEQUENCE DATABASE, 23 June 1997 (1997-06-23), XP002163419 HEIDELBERG DE Accession Nr.: AA478780 abstract	1-42
X	--- STRAUSBERG R ET AL.: "Human cDNA sequence IMAGE:2138166" EMBL SEQUENCE DATABASE, 24 March 1999 (1999-03-24), XP002163420 HEIDELBERG DE Accession Nr.:522149 abstract	1-42
X	--- HILLIER L ET AL.: "Human cDNA clone IMAGE:263887" EMBL SEQUENCE DATABASE, 5 January 1996 (1996-01-05), XP002163421 HEIDELBERG DE Accession Nr.: N28525 abstract	1-42
A	--- "Atlas(tm) human cDNA expression array I" CLONTECHNIQUES, April 1977 (1977-04), pages 4-7, XP002914393 US the whole document	1-20
A	--- REICHERT J ET AL: "HUMAN AND RODENT EXPRESSION PATTERN OF A FUSION GENE ISOLATED FROM AN MCF7 CDNA LIBRARY" INTERNATIONAL JOURNAL OF ONCOLOGY, vol. 9, no. 1, 1996, pages 29-32, XP000906725 page 29	1,6,7,17
A	--- WO 98 40486 A (GENETICS INST) 17 September 1998 (1998-09-17) page 29, line 20 -page 60, line 13 page 18, line 5 -page 26, line 32 -----	1-5, 8-25, 28-46

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 00/01496

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 21-40
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(v) PCT - Presentation of information:
Although claims 21-40 could be considered as a mere presentation of information, according to Rule 39.1(v) PCT, the search has been carried out as far as possible in our systematic documentation.
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-46 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-46, all partially

Invention 1:

A nucleic acid molecule having the sequence of the clone hfbr2_16c16 (corresponding to SEQ.ID.1); an assemblage comprising said nucleic acid; a computer readable medium comprising said nucleic acid; a polypeptide encoded by said nucleic acid; an antibody binding to said polypeptide; an expression vector comprising said nucleic acid and a method for producing said polypeptide.

2. Claims: 1-46, all partially

Invention 2-233:

same as invention 1, but for each single clone as set forth in claim 1 (i.e. starting with clone hfbr2_16f21 and ending with clone hutel_2h3)

NB: for the sake of conciseness, the first subject-matter is explicitly defined, the other subject-matter by analogy thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 00/01496

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0009552	A	24-02-2000	AU	5557099 A	06-03-2000

WO 9840486	A	17-09-1998	US	5976837 A	02-11-1999
			AU	6702298 A	29-09-1998
			EP	0973890 A	26-01-2000

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12659 A2

(51) International Patent Classification⁷: C07K 14/00

[DE/DE]; Grosse Lachstrasse 30a, 69207 Sandhausen (DE).

(21) International Application Number: PCT/IB00/01496

(74) Agent: CARPMAELS & RANSFORD; 43 Bloomsbury Square, London WC1A 2RA (GB).

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,499 18 August 1999 (18.08.1999) US
60/156,503 28 September 1999 (28.09.1999) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/156,503 (CIP)
Filed on 18 August 1999 (18.08.1999)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): GERMAN HUMAN GENOME PROJECT [DE/DE]; Fraunhofer Patentstelle, Leonrodstrasse 68, 80636 Munich (DE).

Published:

— Without international search report and to be republished upon receipt of that report.

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): WIEMANN, Stefan

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN DNA SEQUENCES

(57) Abstract: Novel human cDNA sequence of a clones, the encoded protein sequence of a clones, antibodies and variants thereof, are provided. The disclosed sequence of a clones find application in a number of ways, including use in profiling assays. In this regard, various assemblages of nucleic acids or proteins are provided that are useful in providing large arrays of human material for implementing large-scale screening strategies. The disclosed sequence of a clones may also be used in formulating medicaments, treating various disorders and in certain diagnostic applications.



WO 01/12659 A2

HUMAN DNA SEQUENCES

Background of the Invention

Current methods for testing pharmacological substances rely on a three-stage testing approach to drug development. First, candidate compounds are typically screened in some sort of *in vitro* system, like inhibition of cancer cell growth. Candidates are then tested in an animal model, as a first approximation of systemic effects, including efficacy and toxicity. Compounds that still show promise after these initial *in vivo* screens, finally are tested in humans. Again, human testing typically occurs in three phases: toxicity; preliminary efficacy; and efficacy. The entire process can take more than a decade and cost hundreds of millions of dollars. Aside from the monetary costs and protracted time scale, moreover, current testing regimes waste the lives of countless laboratory animals and needlessly endanger the lives of human subjects.

A need exists, therefore, for more sophisticated drug screening techniques that can be done rapidly *in vitro*. These screening techniques ideally will be reflective of systemic and/or organ-specific responses, so that they provide a reliable indicator of action in a human body. Current techniques, however, tend to utilize only a single or limited number of markers, thus answering only very simple questions that are of questionable medical import. For example, a typical *in vitro* assay may ask whether a lead compound binds a particular receptor, which has been implicated in a certain disorder. It is presumed that such binding is indicative of therapeutic usefulness, but it does not even purport to address systemic effects.

Not only are screening techniques for efficacy inadequate, the available toxicity screens likewise are inadequate. Toxicity, on a first level, is usually measured by animal testing. Aside from the complications related to *in vivo* versus *in vitro* testing, such screens are insufficient because of differences in metabolism, uptake, etc., relative to humans. Thus, improved methods would be not only be *in vitro*-based, they would also be more "human."

With the increasing miniaturization of screening assays and the growing availability of targets for pharmaceutical intervention, there is increasing interest in developing arrays containing large numbers of these targets that can be assayed simultaneously. If such an

array contains a large enough population of targets, it can be used to essentially mimic the systemic response. In other words, the array becomes an *in vitro* surrogate for the human body. The more refined the array, the more accurate the predictive capability. In theory, an array could be constructed that can detect all of the known human expression products simultaneously, thereby, providing a very reliable indicator of the human response to a given compound. These arrays offer advantages over the present *in vitro* screening systems in that they can assay large numbers of responses simultaneously. They are superior to animal testing because they are more "human" and, thus, more predictive of human responses.

In order to construct such arrays, however, the field is in need of further human targets. Advantageously, such targets will be provided with additional physiologically relevant information, such as whether the target is expressed in a particular tissue and whether it is related to a known functional class of targets. In this way, the artisan can focus as needed, for example, on tissue-specific effects or target class-specific effects, thereby providing information useful in evaluating efficacy and/or toxicity.

In addition to a need for pharmacological screening targets, there is a need for further pharmacological substances. These substances can be used in the formulation of medicinal compositions and in treating a wide variety of disorders.

The present invention responds to the aforementioned and other needs in the field by providing a population of novel targets useful, *inter alia*, in the profiling and medicinal contexts described above.

Summary of the Invention

It is an object of the invention, therefore, to provide a set of human cDNA clones. Further to this object, the invention provides sequences of human cDNA clones that were isolated from libraries generated from different human tissues.

It is another object of the invention to provide assemblages of targets useful in profiling matrices for screening pharmacological test compounds. According to this object, assemblages comprising different populations of human nucleic acids, proteins and antibodies are provided. In different embodiments, cDNA library-specific assemblages and target-family-specific targets are provided.

It is a further object of the invention to provide a database of human nucleotide and protein sequences. Further to this object, novel human nucleotide and protein sequences are provided in electronic form. In one embodiment, one or more of these sequences is provided in a searchable database.

It is still another object of the invention to provide biologically active target molecules useful in treating or detecting human disorders. Further to this object, the invention provides nucleic acid and protein molecules that have the capacity to affect disease etiology or symptoms or correlate with known disease states. Also further to this object, a database is provided which comprises the disclosed molecules in electronic form.

It is still a further object of the invention to provide polypeptides encoded by the human cDNA clones disclosed herein. Further to this object, the invention provides antibodies and fragments thereof that are capable of binding to a specific portion of these polypeptides.

It is yet another object of the invention to provide pharmaceutical compositions which comprise an effective amount of a pharmaceutical agent, wherein the pharmaceutical agent is selected from the group consisting of one or more polypeptides contemplated by the invention, variants or functional derivatives thereof, and antibodies thereto; and a physiologically acceptable carrier or excipient.

It is still another object of the invention to provide expression vectors comprising one or more human cDNA clones disclosed herein or fragments thereof; and optionally a promoter operably linked to the cDNA clone or fragment thereof. Further to this object, the invention provides methodology for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by a human cDNA clone disclosed herein.

Detailed Description

The invention results from a need in the art for new human nucleic acids and proteins. This need arises in several contexts. First, there is a need to identify targets for therapeutic intervention. Second, there is a need to identify molecules that may be adversely affected in a therapeutic context, thereby resulting in toxicity. Knowledge of these molecules will aid in

the design of new medicaments with enhanced efficacy and decreased toxicity. Finally, the need encompasses human nucleic acids and proteins that have medicinal applicability in their own right.

In view of these needs, the present inventors set out to isolate and sequence human cDNAs from tissue-specific libraries. In this way, they represent subsets of molecules likely to be targets for therapeutic intervention or for avoiding toxicity. In addition, the inventors divided the molecules into various sub-categories, based on suspected functionality, structural similarity etc, which are of interest from a pharmacological perspective. These molecules are disclosed in provisional application serial nos. 60/149,499 and 60/156,503, filed August 18, 1999, and September 28, 1999, respectively, both of which are hereby incorporated by reference in their entirety.

GENERAL DESCRIPTION OF THE INVENTIVE MOLECULES

The present invention provides novel polynucleotide molecules that, in some instances, have similarities with known molecules. The inventive DNAs were cloned from five different human cDNA libraries. In addition to these DNA molecules, the invention provides their protein translations and antibodies derived from them. The inventive DNA and protein sequences are show individually, below. The inventive nucleic acids also include the complements of these DNA sequences, as well as their RNA counterparts. Methods of producing the molecules also are provided. Further, the invention provides methods for detecting all or part of the molecules and of detecting polynucleotides encoding all or part of the molecules.

The inventive molecules derive from five cDNA libraries: human fetal brain; human fetal kidney; human mammary carcinoma; human testis; and human uterus. For convenience, each sequence bears a designation that indicates from which library it is derived. In particular, these designations are: "hfpbr" for human fetal brain; "hfkf" for human fetal kidney; "hmcfc" for human mammary carcinoma; "htes" for human testis; and "hute" for human uterus. The individual libraries were constructed and screened as described below in the examples.

The protein and DNA molecules of the invention are variously described herein as "target" molecules or "inventive" molecules. The sequences and other information pertinent to the nucleic acid and protein molecules of the invention are shown, below.

Interpreting the data disclosed with the Table and cDNA sequences, below:

The table and data below provide the coding sequences of the inventive cDNAs as well as the protein sequences and other useful information, as set out below.

Grouping

The clones were assigned to the following fourteen functional and/or tissue-derived groups:

1. Cell Cycle
2. Cell Structure and Motility
3. Differentiation/Development
4. Intracellular Transport and Trafficking
5. Metabolism
6. Nucleic Acid Management
7. Signal Transduction
8. Transmembrane Protein
9. Transcription Factors
10. Brain derived
11. Kidney derived
12. Mammary Carcinoma derived
13. Testes derived
14. Uterus derived

Description of Clone Files

The individual clone files are structured in the same pattern. The Sections are separated by paragraphs.

1. Clone Name

The clone names are deciphered with reference to the following example:

DKFZphfkd2_24e23, wherein the code represents:

- producer of library ("DKFZ") (for convenience, this reference may be eliminated)
- a "p" for "plasmid cDNA library" (for convenience, this reference may be eliminated)
- library name (e.g. hfbr = human fetal brain; hfkd = human fetal kidney; hmcf = human mammary carcinoma; htes = human testes; hute = human uterus)
- an underscore (" _ ") to separate library information from plate information
- plate number (e.g. "16")
- plate coordinates (letter first; e.g. "f14")

2. Group

3. Introduction

short review of the similarities, function of the protein and possible applications

4. Short Information

specifications about the cDNA (who sequenced, completeness of the cDNA, similarity, who sequenced, chromosomal localisation, length of cDNA, localisation of poly A tail and polyadenylation signal)

5. cDNA-Sequence**6. BLASTn Results**

search results of blasting the cDNA sequence against all public databases

7. Medline Entries

information about genes/proteins similar to the novel cDNA (if available)

8. Putative Encoded Protein Information

specifications about the encoded protein (ORF: length and localisation of the reading frame)

9. Protein Sequence**10. BLASTp Results**

search results of blasting the protein sequence against all public databases

11. Pedant Information

output of fully automated annotation: summarises peptide information, homologues, patterns as follows:

[Length]

- length of the protein = number of amino acid residues

[MW]

- molecular weight of the protein

[pI]

- isoelectric point

[HOMOL]

- shows protein with closest similarity to the cDNA-encoded protein

[FUNCAT]

- functional information according to a catalogue developed by Munich

Information center for Protein Sequences (MIPS)

[BLOCKS]

- Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The blocks for the Blocks Database are made automatically by looking for the most highly conserved regions in groups of proteins documented in the Prosite Database. The Prosite pattern for a protein group is not used in any way to make the Blocks Database and the pattern may or may not be contained in one of the blocks representing a group. These blocks are then calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of matches. It is these calibrated blocks that make up the Blocks Database. The WWW versions of the Prosite and SWISS-PROT Databases that are used on this server are located at the ExPASy World Wide Web (WWW) Molecular Biology Server of the Geneva University Hospital and the University of Geneva. World Wide Web URL http://blocks.fhcrc.org/blocks/about_blocks.html/ is the entry point to the database.

- here Blocks segments found in the analysed protein sequences are displayed

[SCOP]

Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The scop database provides a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in Brookhaven National Laboratory's Protein Data Bank (PDB). It is available as a set of tightly linked hypertext documents which make the large database comprehensible and accessible. In addition, the hypertext pages offer a panoply of representations of proteins, including links to PDB entries, sequences, references, images and interactive display systems. World Wide Web URL <http://scop.mrc-lmb.cam.ac.uk/scop/> is the

entry point to the database. Existing automatic sequence and structure comparison tools cannot identify all structural and evolutionary relationships between proteins. The scop classification of proteins has been constructed manually by visual inspection and comparison of structures, but with the assistance of tools to make the task manageable and help provide generality. Proteins are classified to reflect both structural and evolutionary relatedness. Many levels exist in the hierarchy, but the principal levels are family, superfamily and fold. The exact position of boundaries between these levels are to some degree subjective. Scop evolutionary classification is generally conservative: where any doubt about relatedness exists, we made new divisions at the family and superfamily levels.

- - here SCOPE segments found in the analysed protein sequences are displayed

[EC]

ENZYME is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and it describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided. World Wide Web URL <http://www.expasy.ch/enzyme/> is the entry point to the database.

- here EC-number and name of enzymes with similarity to the analysed protein sequences are displayed

[PIRKW]

- functional information according to the Protein Information Resource (PIR) database catalogue developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[SUPFAM]

- information according to the Protein Information Resource (PIR) database catalogue of protein superfamilies developed by Munich Information Center for Protein Sequences (MIPS), the National Biomedical Research Foundation (NBRF) and the International Protein Information Database in Japan (JIPID).

[PROSITE]

please refer to 12. PROSITE Motifs

[PFAM]

please refer to 13. PFAM Motifs

[KW]

- overall 2dimensional folding information

- 3D indicates that the proteins is similar to a protein of which a 3 dimensional structure is known

- overall structural information

[]

The last PEDANT-block depicts information about the folding structure of the protein generated by PREDATOR. PREDATOR is a secondary structure prediction program. It takes as input a single protein sequence to be predicted and can optimally use a set of unaligned sequences as additional information to predict the query sequence. The mean prediction accuracy of PREDATOR is 68% for a single sequence and 75% for a set of related sequences. PREDATOR does not use multiple sequence alignment. Instead, it relies on careful pairwise local alignments of the sequences in the set with the query sequence to be predicted.

World Wide Web URL http://www.embl-heidelberg.de/argos/predator/predator_info.html is the entry point to the database.

- H = helix, E = extended or sheet, _ = coil, T = transmembrane, B = beta

- x indicates a low-complexity region with repeat-like structure which is omitted in all BLAST searches

12. PROSITE Motifs

PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs. World Wide Web URL <http://www.expasy.ch/prosite/> is the entry point to the database. A description of the prosite consensus patterns is also provided, below.

13. PFAM Motifs

PFAM (protein families) is a large collection of multiple sequence alignments and hidden

Markov models covering many common protein domains. World Wide Web URL <http://www.sanger.ac.uk/Pfam/> is the entry point to the database.

Deposit of Clones

Clones were deposited as a pool with the American Type Culture Collection under accession number _____, from which each clone comprising a particular polynucleotide is obtainable. Each clone has been transfected into separate bacterial cells (*E. coli*) in this composite deposit.

The clones may also be obtained from the Resource Center of the German Human Genome Project (Heubner Weg 6, 14059 Berlin, GERMANY). The Resource Center library numbers are slightly different than those presented here, but may be readily obtained by the following key or with the assistance of Resource Center personnel.

The library name becomes a number: brain (hfbr2) becomes 564; kidney (hfk2) becomes 566; mammary carcinoma (hmcfl) becomes 727; testis (htes3) becomes 434; and uterus (hute1) becomes 586. Next, the plate number is converted to two digits (e.g., "2" becomes "02") and is moved behind the plate coordinate, and the underscore is dropped. The following examples are helpful:

<u>Listed Number</u>	<u>Resource Center Number</u>
DKFZphfbr2_16f21	DKFZp564F2116
DKFZphfk2_1j9	DKFZp566J091
DKFZphmcf1_1c23	DKFZp727C231
DKFZphtes3_14g5	DKFZp434G0514
DKFZphute1_17k7	DKFZp586K0717

The libraries were constructed using two commercially available vectors. The brain (hfbr2 designations) and kidney (hfk2 designations) libraries utilize pAMP 1 from Life Technologies and are maintained in XL-2Blue (Stratagene); the uterus (hute1), testes (htes3) and mammary carcinoma (hmcfl) libraries are constructed in pSPORT1, also from Life Technologies, and are maintained in DH10B (Life Technologies). In addition to the following techniques, consultation with the commercial literature available on these clones will make evident all of the housekeeping techniques needed to propagate and isolate the individual constructs. All inserts may be excised with a NotI/SalI digestion. Alternatively, universal primers, flanking the cloning region, may be used to amplify the inserts using PCR methods.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. Methods of probe design are presented below.

Oligonucleotide probes may be labeled with γ - ^{32}P ATP (specific activity 6000 Ci/mmole) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other, non-radioactive labeling techniques can also be used. Unincorporated label typically is removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe can be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe generally should be approximately 4×10^6 dmp/pmole.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 50 - 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used). The culture should preferably be grown to saturation at 37°C ., and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 $\mu\text{g/ml}$ (for XL-2Blue strains 25 $\mu\text{g/ml}$ tetracycline should also be used) and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C . Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them. The filter is then preferably incubated at 65°C . for 1 hour with gentle agitation in 6 x SSC (20 x stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 $\mu\text{g/ml}$ of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1×10^6 dpm/mL. The filter is then preferably incubated at 65°C . with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2 x SSC/0.5% SDS at room temperature without agitation, preferably followed by 500 mL of 2 x SSC/0.1% SDS at room

temperature with gentle shaking for 15 minutes. A third wash with 0.1 x SSC/0.5% SDS at 65°C. for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

Alternatively, clones may be grown as described above, and PCR used to isolate the insert DNAs. Methods of PCR are described below and are otherwise well known .

ERROR SCREENING

The DNA sequences found herein derive from individual clones, which are publicly available, as noted above. Thus, the skilled artisan will recognize that any specific sequence disclosed herein readily can be screened for errors by resequencing a particular fragment, in both directions (*i.e.*, by sequencing both strands). Alternatively, error screening can be performed by amplifying and/or cloning any of the inventive DNAs, using for example RT-PCR, and sequencing the resulting amplified product. In the event that there is a sequencing error, reference should be made to the deposited clone as the correct sequence.

USES AND BIOLOGICAL ACTIVITIES OF THE INVENTIVE MOLECULES

The inventive molecules and their derivatives are susceptible to a wide variety of uses, based on functional and/or structural properties. The skilled worker will appreciate, based on the biological activities detailed below, and discussed with regard to the individual sequences disclosed below, that the inventive molecules will find usefulness in numerous therapeutic and diagnostic applications.

The DNA molecules, especially the potassium salts thereof, can be used as fertilizer supplements due to their high nitrogen and phosphorus contents. Since the DNAs are of defined length, they are also useful in gel electrophoresis as molecular weight markers. Due to their similarity with known molecules, certain of the DNA molecules and their variants and derivatives may be used in any number of different diagnostic procedures and therapeutic applications. They may also be used to make the encoded proteins.

The proteins themselves have many possible uses. They may be used as a nutritional supplement for humans, animals and even for laboratory use as, for example, medium for bacterial cultures. Moreover, since the proteins are of defined, known sizes, they may be used as molecular weight markers for gel electrophoresis and gel filtration. Because they are of defined sequences, they also have use in microsequencing and protein fingerprinting applications.

Expression Profiling Applications

Given their known tissue expression and functional associations, assemblages of the inventive proteins (or corresponding antibodies) and nucleic acids are particularly suited to expression profiling applications. Expression profiling generally entails constructing an array of indicators that signal the presence of a particular RNA or protein expression product. Such arrays can be used to evaluate, for example, pharmacological effectiveness and toxicity. In particular, expression profiles from such arrays can be generated from cells treated with known compounds, having known properties, and these profiles can be compared to profiles of unknowns to evaluate similarities and differences, which can be correlated with efficacy or toxicity.

Additional uses of profiling include diagnosis, tracking development, and ascertaining signaling and metabolic pathways. For examples of references describing profiling and its uses, see Farr *et al.*, U.S. Patent 5,811,231 (1998); Seilhamer *et al.*, U.S. Patent 5,840,484 (1998); Rine *et al.*, U.S. Patent No. 5,777,888 (1998); WO 97/27317; WO 99/05323; WO 99/09218; and WO 99/14369. For a device for implementing such techniques, see Lipshutz *et al.*, U.S. Patent No. 5,856,174 (1999) and Anderson *et al.*, U.S. Patent No. 5,922,591 (1999).

In one embodiment, a subset of the inventive DNAs will be arrayed on a substrate, like a gene chip, a filter or a 96-well plate. Test samples containing cells are maintained in the presence of a label capable of incorporation into nascent mRNA. Samples are treated with test and control compounds, which will induce mRNA expression in the sample, resulting in incorporation of label. Whole mRNA is isolated and applied to the array such that it hybridizes with the DNAs contained therein. After washing, the amount of hybridization is quantified and a profile is generated. These steps are repeated with various control and test compounds, thereby generating a library of profiles, which can be used to ascertain the relationships relevant to pharmacological efficacy or toxicity.

The matrices used in such profiling, however, need not be limited to those utilizing DNAs. Rather, other nucleic acids, like RNAs and protein nucleic acids (PNAs), as well as the inventive proteins and antibodies corresponding to the inventive proteins may also be employed. Hence, for example, antibodies could form the array and the samples could be treated in order to label nascent proteins. Whole proteins then would be isolated and applied to the antibody matrix. Developing the resulting signal would result in a protein expression profile, which is useful in essentially the same manner as the nucleic acid profile. A protein matrix could be used, for example, in evaluating antibody responses to pharmaceutical agents in order to eliminate possible cross-reactivity.

Moreover, where nucleic acids are used in the matrix, it is often beneficial to use variants (as defined below) of the molecules described herein. This can be used to account for genetic variations that are of little or no consequence to the function of the resultant gene product. Hence, they can account for wobble or conservative amino acid variations that do not perturb function, like variations in some of the protein motifs elucidated below. Thus, each position in the matrix can employ multiple nucleic acid probes that account for a series of variants.

Expression profiling may also be done, in another embodiment, using two-dimensional protein gels in which the inventive proteins are detected. The resultant profiles can be used in the same way as described.

Matrices useful for profiling may be constructed based on different criteria. Of course, the more relevant profiles will take into account expression of most human genes, preferably all of them. In certain situations, however, it is advantageous to look at a smaller subset. For example, if one were concerned about fetal neural toxicity, a fetal brain-specific matrix might be chosen. On the other hand, if one were interested in targeting mammary carcinoma tissue, a corresponding matrix could be used. Thus, matrices may be constructed using all of the sequences available from a tissue-specific library.

* * *

The following discussion relates to some of the various functional and structural groupings that would be of interest to the artisan wishing to construct profiling matrices. Of course, the artisan will also recognized that these functional descriptions may find additional applicability in the therapeutic and diagnostic applications discussed below.

Cell Cycle

A proliferating cell must coordinate replication and chromosomal separation to ensure that the genome is replicated completely, and that a single copy is correctly inherited by each daughter cell. The cell cycle is the coordinated series of events that achieves these aims. Many of the key events are initiated by a family of conserved Serine/threonine protein kinases, the cyclin-dependent kinases (CDKs), that are activated by the cyclin family of proteins (cyclins A-H). In turn, the cyclin-CDK complexes are modulated by other protein kinases or phosphatases, and by binding specific inhibitor proteins. The enormous variety of ways in which CDK activity can be regulated allows the cell to respond to internal signals generated by preceding events in the cell cycle and to external growth signals.

The somatic cell cycle is divided into four phases: DNA replication (S phase) and chromosome separation (M phase) are separated by gap phases (G1 and G2). At specific control points the decision to begin the next stage (DNA synthesis or mitosis) is carefully regulated.

Cdc2, the primary kinase, is especially required for the G1-S transition and S phase. Cdc4 and Cdc6 are involved at the restriction point, where the cell can decide to proliferate or arrest (G1 \leftrightarrow G0) and Cdc7 is a CDK activating kinase (CAK) as well as a subunit of TFIIH.

The Cyclin-CDK complexes are regulated in various ways. One is through phosphorylation by CDK activating kinases (CAK), like the Y15 kinase (Wee1) and dephosphorylation by CDK associated phosphatases (CAP), like Cdc25A a member of the Cdc25 family (Cdc25A, B and C).

An other way of regulation occurs through two classes of CDK inhibitors (CKI), the INK4 proteins p15, p16, p18, and p19, who negatively regulates the cyclin D CDK complexes and second the p21 family with p21, p27, and p57.

The cell cycle is also regulated through ubiquitin-mediated proteolysis involving the destruction of both cyclins and CDK inhibitors by the 26S proteasome, that requires an ubiquitin conjugating enzyme (UBC) and an ubiquitin ligase. The instability is conferred by PEST regions (cyclin D and E) or a ten amino acid region in the amino terminus (degradation box) in the A- and B-type cyclins.

All these modifications play an important role for the cellular localization, because only the nuclear CDK-cyclin complexes are functional for cell cycle. During G1 phase of the cell cycle, cyclins A, E and D are synthesized and bind to their cyclin-dependent kinase (CDK) partners. CDK complexes containing cyclins A, E and D1 are then imported into and concentrated within nuclei. Cdk6- cyclin D3 has been localized to both cytoplasmic and nuclear compartments, although only the nuclear complex is active. As cells enter S phase, cyclin A and cyclin E complexes remain within the nucleus, whereas cyclin D1 relocates to the cytoplasm for proteolysis at the onset of S phase. Like Cdk2-cyclin A, Cdc2-cyclin A is nuclear and remains so until it is degraded during mitosis. By contrast, as a result of ongoing nuclear import and more rapid re-export, cyclin B1, which binds to Cdc2 upon synthesis during S phase, is predominantly cytoplasmic. Cdc2-cyclin B2 is also cytoplasmic, although this might occur through anchoring of the complex to some cytoplasmic constituent. At prophase, phosphorylation of cyclin B1 promotes accumulation of Cdc2-cyclin B1 in the nucleus, whereas cyclin B2 remains in the cytoplasm until nuclear envelope breakdown.

Two crucial regulators of Cdc2-cyclin B-Wee1 and Cdc25C exist and are responsible for the G2 to M control point. Wee1 is a nuclear protein throughout the cell cycle, whereas Cdc25C binds to 14-3-3 proteins during interphase and remains predominantly cytoplasmic. In some systems Cdc25C, like cyclin B1, rushes precipitously into the nucleus just before entry into mitosis.

The 110-kDa retinoblastoma (tumor suppressor) protein (RB), a pRB-family member is an important regulator of cell-cycle progression and differentiation. Like the E2F family (E2F1-5) or DP family (DP1-3) of transcription activators, RB suppresses inappropriate proliferation by arresting cells in G1 by repressing the transcription of genes required for the transition into S phase. Before the cell proceeds into S phase, RB becomes phosphorylated at multiple sites by the cyclin dependent protein kinases (CDKs) and loses its transcriptional repressing activity. Phosphorylation of RB during late G1 phase results in the dissociation of the E2F-RB repressor complex which allows S-phase specific genes to be transcribed. Cyclin E is the evolutionary conserved target for E2F and interacts together with CDC2 in late G1.

For a proliferating cell it is vital that only undamaged DNA is replicated because if DNA damage is substantial, its replication can lead to chromosome loss or rearrangement.

Thus, we find a G1<->S checkpoint in late G1 that requires tumor suppressor p53. A p53-dependent G1 arrest is effected by the cyclin dependent kinase inhibitor p21 through higher expression levels that inhibits almost all cyclin CDK complexes.

The kinase responsible for phosphorylating the unidentified kinetochore component in metaphase may be a member of the MAP kinase family and appears to be the proto oncogene c-MOS, a cytostatic factor (CSF) in meiosis.

Several categories of proteins are coded for by clones of the invention within the overall group of "Cell cycle" and include, among others, the following:

Tumor suppressors (e.g. N33): Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. The N33 gene has been reported by OMIN OMIN (Online Mendelian Inheritance in Man at <http://www.ncbi.nlm.nih.gov/htbin-post/Omin>) to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) prostate cancer suppression (OMIN *601385). Clones in this category include: fbr2_2k14.

C-TAK1 Cdc25c associated protein kinase: Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. Alterations in the gene coding for the above protein kinase has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Pancreatic cancer (OMIN *60278). Clones in this category include: tes3_7j3.

Cell structure and motility

One of the major differences between prokaryotes and eukaryotes is the ability of the eukaryotic cell to adopt very different shapes dependent on its function during the differentiation process. Animal cells vary from being round to extended cylindric forms like motoneurons or muscle cells. In humans, more than 100 different cell types can be distinguished, each having a characteristic shape. The form of a cell often is closely related to

its capacity to move. Some completely differentiated cells like fibroblasts can still change their form actively, thereby migrating. Other cell types serve as motor elements - “macroscopically” like muscle cells or “microscopically” like ciliated epithelia. Such tasks are fulfilled by a big class of proteins; on the one hand responsible for maintenance of cell structure and contacting neighbor cells or the intercellular matrix and on the other hand for cell motility. These topics cannot be regarded separately: The motility apparatus e.g. must be fixed in the cytoskeleton. Three different types of filaments can be distinguished: Actin filaments, tubulin filaments and intermediate filaments, each present in almost all types of cells.

Actin filaments (F-actin) are built up of monomers (G-Actin). In muscle cells, actin, myosin, for both of which several paralogous genes are known, as well as many more proteins are constituents of the contractile apparatus.

The “thin” and “thick filaments” in a muscle cell consist mainly of actin and myosin, respectively.

Several different proteins are responsible for the anchoring of the actin filaments in the Z-disks (e.g. alpha-actinin and desmin) or at the end of the myofibers in the cell membrane.

Troponin I, -C, -T and Tropomyosin - associated with actin - confer the Ca^{++} -dependent triggering of contraction.

Length of the sarcomere is controlled by the giant protein titin.

In smooth muscle, there is no troponin. Contraction activity is controlled by phosphorylation / dephosphorylation of myosin by a specialized kinase instead. Contractile fibers are not organized in sarcomeres.

Apart from contributing to muscle contraction, the actomyosin system is responsible for many other motions at cellular level, e.g. the amoeboid movement of pseudopodia or the fission of cells at the end of mitosis by a contractile ring.

Besides this, actin fibers fulfill structural tasks like maintenance of the shape of stereocilia or microvilli. Here, actin filaments are connected by proteins like fimbrin. But not

only specialized structures like the mentioned ones contain actin fibers. There is a network covering the complete cell volume with F-actin as a major constituent. Whereas the actin filaments in the structures mentioned above are relatively stable, this F-actin is highly dynamic. Management of the network structure and turnover is achieved by connecting proteins like alpha-actinin, fimbrin or fill-in; turnover is regulated by gelsolin, villin, and different capping- and fragmentation-proteins.

Microtubules are built up of alpha-beta tubulin heterodimers. Turnover of filaments is achieved by building-in and releasing of monomers with different time constant rates at both ends. The resulting cycle is called "treadmilling". Thirteen strings of tubulin duplets build up one subfiber, whereas one fiber contains two or three of those. A complete axoneme consists of 9 radial and 2 central fibers. This "9+2" - structure is the basis both of flagella, their basal bodies and centrioles. In flagella, several additional structures like radial elements exist. Nexin connects the fibers and dyneine is the motor ATPase which shifts the fibers relative to each other. Several genetic diseases like the Cartageneric syndrome are caused by deficiencies of distinct proteins in cilia.

Besides this, microtubules are abundant in all types of cells. They are part of a delivery system for organelles, e.g. in the golgi apparatus. A further very important system based on microtubules is the mitotic spindle, it is organized by the centrosomes. Besides many other components, the major part of a centrosome are two centrioles which are built up of nine microtubule-triplets. Most remarkably, new centrioles are not synthesized de novo but generated by duplication of old ones.

Cytoplasmic microtubules are associated with many different proteins. Two major classes are known: The MAPs ("microtubule-associated proteins", with molecular masses between 200 and 300 kD) and the much smaller tau-Proteins with a MW between 60 and 70 kD. These proteins regulate the treadmill-process and the interaction with other structures in the cell.

Besides actin and myosin the so-called intermediate filaments constitute a third class of filaments. In contrast to the former two groups, they do not participate in motility, nor are they dynamic structures subject to a vivid turnover. The most important ones are

neurofilaments (in neurons), keratin filaments (mainly in epithelial cells), and vimentin filaments (in many sorts different cell types).

The biological function of both the cytoskeleton as well as contractile apparatus of a cell does not end at the cell membrane. Cells must be embedded in the extracellular matrix, all cells of a muscle must act as one single mechanical unit and epithelia must resist macroscopic mechanical forces. Hence, cell adhesion and the extracellular matrix are closely connected to the cytoskeleton. Vincullin is one of the proteins which serve as an anchor for intracellular fibers (actin). Different types of desmosomes and tight junctions connect neighbor cells with intercellular fibers. On the inside, cytoplasmic plaques connect them to the cytoskeleton. These structures, on the one hand, serve as mechanical elements whereas gap junctions, on the other hand, connect cells metabolically.

The extracellular matrix consists of a network of proteins, glycoproteins and polysaccharides. Different proteins are present in relation to different mechanical demands: Elastin is found in tissues with high elasticity (lungs, heart) whereas collagen, a more hard-wearing protein, is found in tendons and ligaments. Fibronectin is an extracellular protein highly important for cell adhesion.

Reference: Murray J *et al* (1992): Cell Motil Cytoskeleton 22: 211-223.

Within the overall group of Cell Structure and Motility several categories of proteins are coded for by clones of the invention:

Collagen alpha chain proteins: Proteins with the typical (xxG)_n repeat of collagen proteins and Pfam von Willebrand factor type A domain(s) suggest they are collagen alpha chains. These proteins can find application in modulation of connective tissue, bone and cartilage development and maintainance. OMIN reports collagen alpha chains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Osteogenesis imperfecta, type I (OMIN #166200); 2) Osteogenesis imperfecta congenita (OMIN #166210); 3) Alport Syndrome, X-linked (OMIN #301050); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Ehlers-Danlos Syndrome, Type VII (OMIN #130060); 6) Marfan Syndrome (OMIN #154700); 7) Alport Syndrome, Autosomal Recessive (OMIN #203780); 8) Alpha-2-Deficient Collagen Disease (OMIN 203760); 9) Goodpasture Syndrome (Omin 233450); 10) Osteogenesis Imperfecta,

progressively deforming, with normal sclerae (OMIN #259420); 11)) Ehlers-Danlos Syndrome, Type VII Autosomal Recessive (OMIN *225410); and 12)) Osteogenesis imperfecta, Type IV (OMIN #166220). OMIN reports that von Willebrand factor type A domains have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases:: 1) Hemophilia A (OMIN *306700); 2) Von Willebrand Disease (OMIN *193400); 3) Giant Platelet Syndrome (OMIN *231200); 4) Thrombastenia of Glanzmann and Naegeli (OMIN *273800); 5) Congenital Thrombotic Diseases due to protein C deficiency (OMIN #176860); 6) Polycystic Kidney Disease 1 (OMIN *601313); 7) Nephrogenic Diabetes Insipidus (OMIN *304800); 8) Factor V Deficiency (OMIN *227400); and 9) Dentatorubral-Pallidoluysian Atrophy (Omin *125370). Clones in this category include: fbr2_2b5.

Radial spokehead protein: Radial spokehead proteins, e.g., Chlamydomonas reinhardtii radial spokehead protein of flagella or axoneme and the Strongylocentrotus purpuratus sea urchin spermatozoa protein p63, and human proteins with similarity thereto are important for the maintenance of a planar form of sperm flagellar beating. The human protein(s) can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men (e.g., in sterility). Clones in this category include: tes3_15i5.

Ankyrins: Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus these proteins are involved in coupling of cyto skeleton and cell membrane. OMIN reports that Ankyrins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Hereditary Spherocytosis (OMIN *182900); 2) Hemolytic Poikilocytic Anemia due to reduced ankyrin binding sites (OMIN 141700); 3) Atypical Elliptocytosis (OMIN 225450); 4) Autosomal recessive spherocytosis (OMIN #270970); 5) Werner Syndrome (OMIN *277700); and 6) Rhesus-unlinked type Elliptocytosis (OMIN #130600). Clones in this category include: tes3_1817.

FGD1-related F-actin binding protein (Farbin/FGD1): FGD1-related F-actin-binding protein (Farbin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. (OMIN 305400). Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as

described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. Clones in this category include: tes3_72k15.

Paramyosins: Paramyosin is a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. Clones in this category include: tes3_7b22.

Tuftelin: Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix. The new protein can find application in modulation of tissue-calcification, especially the uterus. As reported by OMIN, tuftelin has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with amelogenesis imperfecta (OMIN *600087). Clones in this category include: ute1_19g22.

Cell Adhesion Regulator (CAR1): CAR1 is involved in the regulation of cell-cell adhesion. OMIN reports the association (as potentially diagnostic, therapeutic, causative, and/or related, etc...) of CAR1 with tumor suppression by the reduction of tumor invasion (OMIN *116935). Clones in this category include: ute1_24j6.

Differentiation/Development

Almost every multicellular organism originates from meiotic cell divisions and the recombination of a paternal and a maternal set of chromosomes. After fertilization of the egg, all cells of a body originate from this one cell. Thus the cells of the developing body are initially genetically alike. But phenotypically they become very different. They are specialized to a certain cell type and arranged in an organized pattern to a certain type of tissue and the whole structure has the well-defined shape of an organ. All these features are determined by the DNA sequence of the genome, which is reproduced in every cell. Each cell acts on the genetic instructions given to a certain time and at a certain place of development and plays its individual part in the multicellular organism. Cell differentiation may be divided into three general steps: cell cycle exit, apoptosis protection and tissue specific gene

expression. These processes are coordinated to provide the final and unique tissue characteristics.

An animal cell that has achieved a certain level of development is said to be determined. This differentiation of a cell may be irreversible and in that case the cell may be renewed only by simple duplication. Other cells are renewed by means of stem cells which are immortal (e.g. stem cells of the bone marrow, epidermal stem cells). The genetic control of development is extensively studied in non-vertebrates and vertebrates. The classical animal model is the fruit fly *Drosophila* and the modern model is the transgenic mouse. Animal transgenesis has proven to be useful for physiological as well as physiopathological studies. Besides the approach based on the random integration of a DNA construct in the mouse genome, gene targeting can be achieved using totipotent embryonic stem cells for targeted transgenesis. Transgenic mice are then derived from the embryonic stem cells. This allows the introduction of null mutations in the genome (so-called knock-out) or the control of the transgene expression by the endogenous regulatory sequence of the gene of interest (so-called knock-in). Mice can be created that express wild-type genes, mutant genes, marker genes or cell lethal genes in a tissue specific manner. These animal models allow to follow changes in tissue and organ development and lead to a better understanding of the cellular function of many genes or to the generation of animal models for human diseases. Fundamental problems in immunology, onset and development of cancer, regulation in fatty acid metabolism, aspects of cardiovascular function, control of the central nervous system development, analysis of reproductive development and function are only some examples of research interests.

The final stage of cell differentiation is growth arrest. In animal tissues with rapid cell turnover terminally differentiated cells undergo programmed cell death. The cells have the ability to kill themselves by activating an intrinsic cell suicide program when they are no longer needed or have become seriously damaged. The execution of this program is termed apoptosis. Apoptosis is of importance for development and homeostasis of animals. The key components of this program have been conserved in evolution from worms (*C. elegans*) to insects (*Drosophila*) to humans. The roles of apoptosis include the sculpting of structures during development, deletion of unneeded cells and tissues, regulation of growth and cell number, and the elimination of abnormal and potentially dangerous cells. In this way

apoptosis provides "quality control mechanism" that limits the accumulation of harmful cells, such as virus-infected cells and tumor cells. On the other hand inappropriate apoptosis is associated with a wide variety of diseases, including AIDS, neuro-degenerative disorders and ischemic stroke. Because it is now clear that apoptosis is a result of an active, gene-directed process, it should be eventually possible to manipulate this form of cell death by developing drugs that interact with its recently identified mechanisms of action. Inducers of cell differentiation, cell cycle arrest and apoptosis might be the novel molecular targets for new anticancer agents in addition to the signaling pathways for growth factors and cytokines.

Proteins, factors, receptors and genes of importance in apoptosis:

Proteases:

- Calpain, an intracellular cysteine protease, exact role unknown.
- Caspase-1 to Caspase-11, a family of proteases synthesized as an inactive proenzyme. Targets of the activated enzymes include: poly(ADP-ribose) polymerase, DNA-dependent protein kinase, U1 ribonucleoprotein, nuclear laminins and cytoskeleton components (actin).
- Granzyme B, a serine protease released by cytotoxic T-cells.

Receptors:

- CD 95 (synonyms: Fas, APO-1), a receptor protein of the TNF-receptor family which includes TNF-R1 and TNF-R2 with the common characteristic of a 70 amino acid cytoplasmic domain.
- FADD (synonym: MORT-1), a cytoplasmic protein
- DR-3 (synonym: APO-3) a member of the TNF-receptor-family
- DR-4 and DR-5

Genes:

- ced-3, ced-4 and ced-9 encode the general apoptotic and antiapoptotic program in *Caenorhabditis elegans*. Apaf-3 is the mammalian homologue of ced-3.

- Bcl-2 / Bcl-xL / Bax / Bcl-xS / Bak: a large gene family that can either inhibit or promote apoptosis.

- Cytokine response modifier A, a cowpox virus gene whose gene product inhibits caspases.

Others:

- Caspase-activated DNase (CAD) and its inhibitor (ICAD), causes DNA fragmentation in the nucleus

- Ceramide, a complex lipid that acts as a second messenger.

- c-Jun N-terminal kinase (JNK) is a proline-directed kinase

- p53 protein, is essential for the induction of apoptosis as a response to chromosomal damage.

- RAIDD, a death signal-transducing protein.

- Receptor interacting protein (RIP) is an accessory protein with a death domain and a serine/threonine kinase activity.

- Sphingomyelinase, an enzyme that hydrolyzes the complex lipid sphingomyelin to ceramide.

- Tumor necrosis factor (TNF) is a type -II membrane protein

- TNF-receptor associated factor (TRAF2), is an accessory protein that can bind to both TNF-R1 and TNF-R2.

Within the overall group of Differentiation/Development, several categories of proteins are coded for by clones of the invention:

Interleukins (e.g. Interleukin-7): Interleukin precursors related to interleukin-7, for example, are expected to act as new growth factors for human B lineage cells. Additionally,

these proteins should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. These interleukins could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells. (OMIN *146660). Clones in this category include: tes3_35e21.

Testis-specific Y-encoded proteins: The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. Proteins of the TSPY-SET-NAP1L1 family represent proteins closely related to TSPY. These proteins seem to be involved in early spermatogenesis. Clones in this category include: fbr2_2d15.

Intracellular transport and trafficking

Eukaryotic cells rely for their viability on the partitioning of many basic cellular processes into membrane-bounded organelles. These are the nucleus, endoplasmic reticulum (ER), Golgi apparatus, endosomes, lysosomal compartments, mitochondria and peroxisomes. Most molecules destined for the lysosome, cell surface and outside the cell are routed through the ER and Golgi, which together with the vesicular intermediates between them, comprise the secretory pathway (Palade 1975). In the ER and Golgi compartments proteins are sorted, modified and often assembled into complexes *en route* to their final destination. Incorrectly assembled proteins are retained in the ER until they fold correctly or are targeted for degradation. Additional proteins are translocated into and function within the luminal spaces of organelles or are secreted. Thus a large proportion of proteins synthesized require targeting to membranes either for insertion into or transport across them. A major purpose of this is growth. The secretory pathway is dependent on an intact cytoskeleton and also closely linked to general metabolism by affecting ribosome biogenesis (Mizuta and Warner, 1994). A huge number of proteins is required for targeting, translocation and sorting of newly synthesized proteins.

The first step in sorting is the recognition of cis-acting targeting or signal sequences that organelle-targeted proteins contain. This is carried out by cytosolic targeting factors and/or receptors on the membrane to which the protein is targeted. In some cases the primary

sequences are extremely degenerate, with only the overall character being conserved (hydrophobicity for an ER signal sequence, helical amphiphilicity for mitochondrial targeting sequence (Kaiser *et al.*, 1987; Lemire *et al.*, 1989). Following the targeting step, proteins are either inserted into or transported across the membrane (translocated) through a proteinaceous apparatus (termed the translocon). The translocon include or recruit motors to drive the translocation process in the correct direction (Schatz and Dobberstein, 1996).

Defined intracellular protein transport steps:

- ER
 - targeting to the ER
 - translocation into the lumen of the ER, and, depending on the presence of certain signals in the peptide sequence transport through the golgi complex
- Mitochondria
 - targeting
 - translocation
- Peroxisomes
- The general secretory pathway
 - protein modification, assembly and quality control in the ER
 - vesicle-mediated trafficking
 - vesicle docking and fusion
 - transport through the golgi apparatus and sorting at the trans-golgi
 - transport to the cell surface
 - transport routes to the lysosome
- Endocytosis
- Specialized protein transport routes
- Protein export from the cytoplasm

References: Palade, G (1975) Science 189:347-358; Mizuta et al. (1994) Mol Cell Biol 14: 2493-2502; Kaiser *et al.* (1987) Science 235: 312-317; Lemire *et al.* (1989) J Biol Chem 264: 20206-20215; Schatz et al. (1996) Science 271: 1519-1526.

Rab proteins

In eukaryotic cells the compartmentalisation of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins

and other molecules. Trafficking between organelles within the secretory pathway occurs as vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated α -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation

inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle, most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown *in vitro* to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca²⁺-binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca²⁺ influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A

homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991) *Mol. Cell. Biol.* 11, 872-885; Echard et al. (1998). *Science*. 279, 580-585; Geppert et al. (1998) *Annu. Rev. Neurosci.* 21, 75-95; Guo et al. (1999). *EMBO J.* 18, 1071-1080; Kato et al. (1996) *J. Biol. Chem.* 271, 31775-31778; Novick et al. (1997) *Curr. Opin. Cell Biol.* 9, 496-504; Peterson (1999) *Curr. Biol.* 9, 159-162; Poirier et al. (1998) *Nat. Struct. Biol.* 5, 765-769; Vitale et al. (1998) *EMBO J.* 17, 1941-1951; Wang et al. (1997) *Nature*. 388, 593-598; Yang et al. (1999) *J. Biol. Chem.* 274, 5649-5653.

Within the overall group of Intracellular Transport and Trafficking several categories of proteins are coded for by clones of the invention.

Rab proteins:

Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells. . Clones in this category include: fbr2_2i17, fbr2_3b16.

Rab10 appear concentrated on membranes in the perinuclear region. Rab 10 has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Choroideremia (OMIN *303199); and 2) RETT Syndrome (OMIN 312750). Clones in this category include: fbr2_62l19.

In mice, Rab17 shows epithelial cell specificity. Rab 17 is discussed as candidate gene for the mouse mutations In (leaden), Tw (twirler), and ax (ataxia). Cloned from a brain cDNA library, the new putative Rab-protein is expected to be involved in vesicle trafficking within neuronal cells. These proteins can find application in modulating the transport of vesicles inside neuronal cells, which are essential for development of functional dendritic processes. . . Clones in this category include: fbr2_41m15.

Ankyrin G: The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments. Ankyrin G has been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with Werner disease (OMIN *277700). Clones in this category include: fkd2_24p5.

Zn-T-transporters: The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide. These proteins can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation. (OMIN *602878, *602095). Clones in this category include: fbr2_62f10.

Metabolism

This group includes proteins which are involved in the uptake and consumption of nutrients, and enzymes which are part of the biochemical pathways for energy metabolism or

which are involved in the supply of building blocks of nucleic acids, proteins (NTPs, dNTPs, amino acids) for DNA/RNA and protein synthesis, and fatty acids (membranes), to allow for the generation of higher order structures. This group constitutes the most important and largest group in prokaryotes and lower eukaryotes. The higher the evolutionary level of an organism is, however, the more other protein classes like 'signal transduction', 'cell cycle' and 'differentiation and development' increase in importance and number of representatives.

Proteins involved in the metabolism of energy and compounds (here: other than nucleic acids or proteins) are usually the products of house keeping genes, they are often constitutively and/or ubiquitously expressed.

Several categories of proteins are coded for by clones of the invention within the overall group of Metabolism:

NAT1, ARD1: In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. These can find application modulating NAT assembly and action and therefore could be important in metabolism of drugs and environmental mutagens.(OMIN *108345). Clones in this category include: fbr2_3g8.

Apolipoprotein E receptor: In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. These proteins can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins. In normal individuals, chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. In familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants because of a defect in apolipoprotein E. Accumulation of the remnants can result in xanthomatosis and premature coronary and/or peripheral vascular disease. OMIN reports that apolipoprotein has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Familial combined hyperlipidemia (OMIN 144250); and 3) Alzheimer disease. (OMIN #104300). Clones in this category include: fbr2_62017.

Ubiquitin carboxyl-terminal hydrolases: Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. OMIN reports that Ubiquitin-specific proteases have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Lung carcinoma (OMIN *603486); 2) x-linked retinal diseases (OMIN *300050); 3) oncogenesis (OMIN *300050); 4) ovarian cancer (OMIN *300050). Clones in this category include: fbr2_78k24; htes3_27d1.

Phosphoserine signature (phosphoglucomutases, phosphomannomutase): These proteins take part in the conversion of hexose phosphates. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Fanconi-Bickel Syndrome (OMIN #227810). Clones in this category include: fkd2_24b15.

NADH ubiquinone oxidoreductase: NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following disease: Brancio-oto-renal syndrome (OMIN *6601445). Clones in this category include: fkd2_3o17.

Transketolases: Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: Wernicke-Korsakoff Syndrome (OMIN *277730). Clones in this category include: tes3_17117.

Fatty acid-CoA synthetases/ligases: These proteins contain AMP-binding domain signature(s), which is present in enzymes which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic,

causative, and/or related, etc...) with the following diseases: 1) Alport syndrome , mental retardation and elliptocytosis (OMIN *300157); 2) Adrenoleukodystrophy (OMIN *300100). Clones in this category include: tes3_35k17.

ADP/ATP or Adenine Nucleotide Translocators: These proteins contain mitochondrial energy transfer signature(s) and are most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.. OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) cardiomyopathy (OMIN *103220); 2) myopathy (OMIN *103220); 3) Progressive external ophthalmoplegia (OMIN *601227). Clones in this category include: tes3_35n12.

Carboxylesterases: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hepatic carboxylesterase with detoxification of foreign compounds (OMIN *114835); 2) non-Hodgkin lymphoma (OMIN *114835); 3) B-cell chronic lymphocytic leukemia (OMIN *114835); 4) rheumatoid arthritis (OMIN *114835). Clones in this category include: tes3_35n9.

Heat shock proteins: OMIN reports that these proteins have associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) 27 kd heat shock protein has been correlated with thermotolerance in response to environmental challenges and developmental transitions. (OMIN *6021295). Clones in this category include: utell_23e13.

Nucleic acid management

The genetic information is stored in the form of nucleic acids in all organisms. Two kinds of nucleic acids exist, DNA and RNA. Whereas the more stable DNA in most organisms constitutes the storage form of the genetic information, the labile RNA and in particular mRNA is an intermediate used for the temporal expression of specific genes.

In eukaryotes, DNA is usually a double stranded linear molecule consisting of two antiparallel strands and made up of a deoxyribose, a phosphorus backbone and the four bases A, C, G, and T. The DNA of some organisms has a ring structure. The structure of DNA was

unraveled years ago by Watson and Crick. DNA is directional molecule determined by the C-atoms of the sugar.

The most important processes dealing with nucleic acids are:

- replication (e.g. DNA polymerases, Telomerase)
- transcription (RNA polymerases)
- RNA processing (maturation - splicing and degradation)
- in addition, enzymes and proteins exist which require a nucleic acid (mostly RNA) in the active center to be functional (ribozymes - e.g. RNase, Ribosomal proteins)

The DNA of a cell is replicated in the S-phase of the cell cycle. Several enzymes carry out the task of doubling this nucleic acid. As all steps of the cell cycle, also the process of replication is tightly regulated. The enzyme DNA polymerase and several other proteins are involved in this process. Whereas many prokaryotes do have only one origin of replication (i.e., the starting point of the replication cycle), in eukaryotic DNAs (chromosomes) multiple such start points exist. The switch from the synthesis (S) phase to the subsequent G2 or M phases of the cell cycle are dependent on the completion of the replication. This makes clear, that a number of proteins are involved in the replication itself as well as in the control of the process. Since most eukaryotic chromosomes are linear structures, additional proteins and enzymes are necessary to make sure that the structure is maintained through successive generations. This includes those proteins necessary to build the three dimensional structure of chromosomes (e.g. histones) and the structural network of the nucleus and nucleolus (including the defined localization of transcriptionally active genes in the vicinity of nucleoli) but also such enzymes as telomerase which guarantees the integrity of the chromosomal ends.

The expression of genes is usually performed in two steps. First a messenger RNA (mRNA) is produced (transcribed) in one to many copies and second this mRNA is translated into the protein product. The regulation of transcription is discussed under the separate heading 'transcription factors', but also the classes 'signal transduction', 'development', 'cell cycle' and others are affected as the expression of certain genes determines the fate of a cell or organism.

The primary transcript (hnRNA - heterogeneous nuclear RNA) is a single stranded one-to-one copy of the gene as it is located on the chromosome. Before a protein can be translated, already during transcription the process of maturation is initiated. Firstly, a 5' cap structure is enzymatically and covalently added to the RNA, blocking the 5' end of the RNA.

Second, when the RNA polymerase has terminated polymerization, the enzyme poly A polymerase adds varying numbers of adenine residues to the 3' end of the transcript. This enzyme recognizes the sequence AAUAAA or AUUAAA (+ some minor variations), cuts the RNA 10 - 30 nucleotides downstream and adds the A residues. The size of the poly A sequence affects the stability of the RNA. Finally, in the process of splicing, the introns present on the genomic level and also present in the hnRNA are spliced out by a multi-protein complex consisting of several proteins and RNAs. The finally matured mRNA is exported to the cytoplasm where it is translated with help of the ribozymes.

The half life of RNA is usually much shorter than that of DNA. Usually, the mRNA is degraded shortly after synthesis, to guarantee a very defined window of expression of a given gene. This regulation is necessary to specifically maintain or change the set of proteins present at any time in a cell. Specific regions in the 3'UTR (untranslated region) determine the stability of the mRNA in the cytoplasm before it is degraded by RNases, enzymes consisting both of protein and RNA.

References: Watson and Crick (1953) *Nature* 171: 737-738.

Several categories of proteins are coded for by clones of the invention within the overall group of "Nucleic acid management" and include, among others, the following:

RNA helicases including DEAD/H box helicases: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. DEAD box proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by with the following disease processes and/or genes: 1) ataxia-telangiectasia gene: "A human gene (DDX10) encoding a putative DEAD-box RNA helicase at 11q22-q23" *Genomics* 33:199-206, 1996, Savitsky et al., (OMIN *601235); 2) hematopoietic tumors: "Cloning and expression of a murine cDNA homologous to the human RCK/P54, a lymphoma-linked chromosomal breakpoint 11q23", *Gene* 166:293-6, 1995, Seto et al. (OMIN *600326); 3) dermatomyositis: a) "The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation."

Arthritis Rheum. 38: 1389-1399, 1995, Seelig et al. (OMIN *603277); b) "Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen." (Letter), *Arthritis Rheum.* 39: 1769-1771, 1996., Seelig et al. (OMIN *603277); c) "The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities", *Cell* 95: 279-289, 1998. Zhang et al. (OMIN *603277); 4) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200); 5) Mucopolysaccharidosis Type IVA (OMIN *253000); 6) Albinism I (OMIN *203100); 7) Wilms Tumor 1 (OMIN *194070); 8) Spinocerebellar Ataxia 7 (OMIN *164500). Clones in this category include: fbr2_23b10, fbr2_3cl8, fbr2_6o17, fbr2_82i24, and tes3_14h21.

Inorganic pyrophosphatase: Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Clones in this category include: fbr2_64a15.

DNA-damage –inducible protein (dinP) or Proteins induced by DNA-Damage: The dinB/P pathway is a second SOS-pathway in E.coli. Genes related to this seem to be involved in modulating DNA repair and mutagenesis. Clones in this category include: fbr2_72b18.

Proteins with myc-type, helix-loop-helix dimerization domain signature(s). This helix-loop-helix domain mediates protein dimerization has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, these proteins could be novel DNA-binding proteins. Clones in this category include: fbr2_72i12.

Cytosolic ribosomal proteins L36: L36 seems to be part of the eukaryotic ribosomal peptidyl transferase center and can find application in modulation of ribosome assembly, maintenance and activity. Clones in this category include: fkd2_3b2.

Ribonuclease H: Ribonuclease H proteins are RNA modifying proteins and have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases as reported by OMIN: 1) Adenomatous Polyposis of the Colon (OMIN

*175100); 2) Retinoblastoma (OMIN *180200) ; and 3) Von Hippel-Lindau Syndrome (OMIN *193300). Clones in this category include: phtes3_15j3.

Signal transduction

Cells in higher order organisms need to continuously communicate with its environment especially with other cells of the same organism in order to maintain the function and specialization of the whole system these cells are part of. This important task of communication is performed with help of cell-surface receptors which receive and transmit signals from outside into the cell.

G-proteins

The largest known family of cell-surface receptors is that of the G-protein-coupled receptors, which mediate the transmission of diverse stimuli such as neurotransmitters, glycopeptides, hormones, peptides, odorant molecules, and photons. The functional unit of these receptors is composed of the receptor molecule itself (GPCR) which is anchored in the cytoplasmic membrane with seven membrane spanning domains, the heterotrimeric G-protein which is composed of α and $\beta\gamma$ -subunits ($G\alpha$ and $G\beta\gamma$), and the effectors that interact with $G\alpha$ and / or $G\beta\gamma$. In particular, the dissociated $G\alpha$ and $G\beta\gamma$ can regulate the activities of a number of effector molecules such as adenylate cyclases, phospholipase C isoforms, ion channels, and tyrosine kinases, resulting in a variety of cellular functions. The process of signal transduction must be tightly regulated and reversible in order to avoid overstimulation, to achieve signal termination, and render the receptor responsive to subsequent stimuli [Iacovelly L. et al., (1999) *FASEB J.* 13, 1-8, Hamm, H.E. (1998) *J. Biol. Chem.* 273, 669-672].

G-proteins are GTPases that, upon binding of GTP change their conformation which in return unmask structural motives, in particular the so called effector loop, which can mediate the interactions to target proteins, or effectors, for the GTPases. This ability enables the GTPases to cycle between active, GTP-bound and inactive, GDP bound conformations and in the process to function as molecular traffic lights in a multitude of signal transduction pathways. The most important of these signal transduction pathways that are regulated with help of G-proteins are that of the phospholipase C / protein kinase C and that of the adenylate cyclase / protein kinase A.

The cycling of GTPases is tightly regulated by three main classes of proteins: The exchange of hydrolyzed GDP for a fresh GTP is facilitated by guanosine nucleotide exchange factors (GEFs), the hydrolysis of GTP to GDP is sped up by GTPase-activating proteins (GAPs), and the dissociation of GDP from the GTPases is inhibited by GDP dissociation inhibitors (GDIs) [Tapon and Hall (1997) *Curr. Opin. Cell. Biol.* **9**, 86-92, Van Aelst and D-Souza-Schorey (1997) *Genes Dev.* **11**, 2295-2322].

SOC-family

A conserved motif that was originally identified in proteins that negatively regulate the signaling action of cytokines was termed SOCS box, the Suppressor Of Cytokine Signaling. Based on homology, five distinct structural protein classes have been identified since that carry this motif. The function of most of these proteins is presently not known. Common to the proteins is only the SOCS box which is located near the C-terminus of the respective peptides. Recently, the SOCS box has been demonstrated to induce binding of proteins to elongins B and C which could target the proteins (and bound substrates) to the proteasomal protein degradation pathway (Kamura, T. *et al.* (1998) *Genes Dev.* **12**, 3872-3881; Zhang, J.-G. *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2071-2076).

The class where the SOCS box was originally described contains several members (SOCS-1-SOCS-7 and CIS). In addition to the SOCS box, these proteins also contain a SH2 (Src-homology 2) domain and a variable N-terminus. These SOCS proteins appear to form part of a classical negative feedback loop that regulates cytokine signal transduction. Upon cytokine stimulation, expression of SOCS proteins is rapidly induced and the proteins inhibit further cytokine action. The mode of action of the SOCS proteins is variable. While SOCS-1 binds and inhibits the JAK (Janus kinases) family of cytoplasmic protein kinases [Narahzaki M. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13130-13134, Nicholson, S.E. *et al.* (1999) *EMBO J.* **18**, 375-385], CIS appears to act by competing with signaling molecules such as the STATs (Transducers and Activators of Transcription) family for binding to phosphorylated receptor cytoplasmic domains [Yoshimura, A. *et al.* (1995) *EMBO J.* **14**, 2816-2826; Matsumoto, A. *et al.* (1997) *Blood* **89**, 3148-3154].

A second class of SOCS box protein contains additionally WD-40 repeats which were initially identified in the mouse WSB-1 and -2 proteins. The functions of WD-40 proteins are not completely understood but seem to be rather divergent. In Cdc4p the WD-40 repeats probably are necessary for binding the substrate for Cdc34p [Mathias, N. *et al.* (1999) *Mol.*

Cell Biol. **19**, 1759-1767]. Cdc4p is a component of a ubiquitin ligase that tethers the ubiquitin-conjugating enzyme Cdc34p to its substrates. The posttranslational modification of a protein by ubiquitin usually results in rapid degradation of the ubiquitinated protein by the proteasome. The transfer of ubiquitin to substrate is a multistep process where WD-40 repeats might play an important function.

Other WD-40 containing proteins (e.g. the retino blastoma binding protein RbAp48) have been shown to bind metal ions (Zinc) and that this metal binding might mediate and/or regulate protein-protein interactions which are functionally important in chromatin metabolism [Kenzior, A.L. and Folk, W.R. (1998) *FEBS Lett.* **440**, 425-429]. These proteins are involved in the RAS-cAMP pathway that regulates cellular growth [Ach R.A. *et al.* (1997) *Plant Cell* **9**, 1595-1606].

The SPRY domain has been identified in pyrin or marenostin, a protein which is mutated in patients with Mediterranean fever and which is similar to the butyrophilin family. While butyrophilins seem to be involved in the lactation process in mammals, the function pyrin is unknown. Three proteins (SSB-1 to -3) have been identified to contain both SPRY and SOCS box motifs. The function of these proteins is also not known.

Ankyrin repeat containing proteins share a 33-residue repeating motif, an L-shaped structure with protruding β -hairpin tips which mediate specific macromolecular interactions with cytoskeletal, membrane, and regulatory proteins. These proteins play fundamental roles in diverse biological activities including growth and development, intracellular protein trafficking, the establishment and maintenance of cellular polarity, cell adhesion signal transduction, and mRNA transcription. Three proteins that contain ankyrin repeats (ASB-1 to -3) have been identified to contain a C-terminal SOCS box additionally to the ankyrin repeats. The function of these proteins or the individual domains remains to be discovered [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

A few small GTPases (RAR and RAR like) do also contain a SOCS box. GTPases are involved in signal transduction during cellular communication. The function of the SOCS box in this type of proteins is currently unclear [Hilton, D.J. *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**, 114-119].

Ca²⁺ as second messenger

The bivalent cation Ca²⁺ is, besides cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very

low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

cAMP

The cyclic AMP is produced by the enzyme adenylate cyclase in response to extracellular signals. Certain G-proteins stimulate the activity of adenylate cyclase which converts ATP to cAMP and PPi. Two molecules of cAMP bind to each of two regulatory subunits of cAMP dependent protein kinase which in turn dissociate from the two catalytic subunits of the heterotetramer R_2C_2 . Upon release of the C-subunits, they become active and phosphorylate substrate proteins at Ser and Thr residues. The process leading from binding of extracellular molecules to their receptors, the transmission of the stimuli into the cell, the activation of adenylate cyclase and the subsequent activation of cAMP dependent protein kinase is one of two major signal transduction pathways in eukaryotic cells. Since the phosphorylation of proteins is a posttranslational modification of proteins, the kinases are described in the class "signal transduction."

SARA

Members of the transforming growth factor β (TGF β) superfamily signal through a family of cell-surface transmembrane serine/threonine kinases, known as type I and type II receptors (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Ligand induces formation of heteromeric complexes of these receptors, and signaling is initiated when receptor I is phosphorylated and activated by the constitutively active kinase of receptor II (Wrana et al., 1994). The activated type I receptor kinase then propagates the signal to a family of intracellular signaling mediators known as Smads (contraction of the *C.elegans* Sma and *Drosophila* Mad genes which were the first identified members of this class of signaling effectors).

Three classes of Smads with distinct functions have been defined: the receptor-regulated Smads, which include Smad1, 2, 3, 5, and 8; the common mediator Smad, Smad4; and the antagonistic Smads, which include Smad6 and 7 (Heldin et al., 1997; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). Receptor-regulated Smads (R-Smads) act as direct substrates of specific type I receptors, and the proteins are phosphorylated on the last two serines at the carboxyl terminus within a highly conserved SSXS motif (Macías-Silva et al., 1996 ; Abdollah et al., 1997 ; Kretzschmar et al., 1997 ; Liu et al., 1997b ; Souchelnytskyi et al., 1997). Regulation of R-Smads by the receptor kinase provides an important level of specificity in this system. Thus, Smad2 and Smad3 are substrates of TGF β or activin receptors and mediate signaling by these ligands (Macías-Silva et al., 1996 ; Liu et al., 1997b ; Nakao et al., 1997), whereas Smad1, 5, and 8 are targets of BMP receptors and propagate BMP signals (Hoodless et al., 1996 ; Chen et al., 1997b ; Kretzschmar et al., 1997 ; Nishimura et al., 1998). Once phosphorylated, R-Smads associate with the common Smad, Smad4 (Lagna et al., 1996 ; Zhang et al., 1997), and mediate nuclear translocation of the heteromeric complex. In the nucleus, Smad complexes then activate specific genes through cooperative interactions with DNA and other DNA-binding proteins such as FAST1, FAST2, and Fos/Jun (Chen et al., 1996 , Chen et al., 1997a ; Liu et al., 1997a ; Labbé et al., 1998 ; Zhang et al., 1998 ; Zhou et al., 1998). In contrast to R-Smads and Smad4, the antagonistic Smads, Smad6 and 7, appear to function by blocking ligand-dependent signaling (reviewed in Heldin et al., 1997).

Phosphorylation of R-Smads by the type I receptor is essential for activating the TGF β signaling pathway (Heldin et al., 1997 ; Attisano and Wrana, 1998 ; Kretzschmar and Massagué, 1998). However, little is known of how Smad interaction with receptors is controlled. A novel Smad2/Smad3 interacting protein has been described (Tsukazaki T. et al., 1998) that contains a double zinc finger, or FYVE domain, and which has been called SARA (Smad anchor for receptor activation). The SARA motif recruits Smad2 into distinct subcellular domains and co-localizes and interacts with TGF β receptors. TGF β signaling induces dissociation of Smad2 from SARA with concomitant formation of Smad2/Smad4 complexes and nuclear translocation. Moreover, deletion of the FYVE domain in SARA causes mislocalization of Smad2 and inhibits TGF β -dependent transcriptional responses. Thus, SARA defines a component of TGF β signaling that functions to recruit Smad2 to the receptor by controlling the subcellular localization of Smad.

References: Abdollah et al. (1997) *J. Biol. Chem.* 272, 27678-27685; Attisano et al. (1998) *Curr. Opin. Cell Biol.* 10, 188-194; Chen et al. (1996) *Nature* 383, 691-696; Chen et al. (1997a) *Nature* 389, 85-89; Chen et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 12938-12943; Heldin et al. (1997) *Nature* 390, 465-471; Hoodless et al. (1996) *Cell* 85, 489-500; Kretschmar et al. (1998) *Curr. Opin. Genet. Dev.* 8, 103-111; Kretschmar et al. (1997) *Genes Dev.* 11, 984-995; Labbé et al. (1998) *Mol. Cell* 2, 109-120; Lagna et al. (1996) *Nature* 383, 832-836; Liu et al. (1997a) *Genes Dev.* 11, 3157-3167; Liu et al. (1997b) *Proc. Natl. Acad. Sci. USA* 94, 10669-10764; Macías-Silva et al. (1996) *Cell* 87, 1215-1224; Nakao et al. (1997) *EMBO J.* 16, 5353-5362; Nishimura et al. (1998) *J. Biol. Chem.* 273, 1872-1879; Souchelnytskyi et al. (1997) *J. Biol. Chem.* 272, 28107-28115; Tsukazaki et al. (1998) *Cell* 95, 779-791; Wrana et al. (1994) *Nature* 370, 341-347; Zhang et al. (1997) *Curr. Biol.* 7, 270-276; Zhang et al. (1998) *Nature* 394, 909-913; Zhou et al. (1998) *Mol. Cell* 2, 121-127.

Calcium

The bivalent cation Ca^{2+} is, along with cAMP, one of the two major second messengers in eukaryotic cells. Its intracellular concentration is tightly regulated and usually kept very low compared to the cell's environment. Ca^{2+} binding proteins and transporters (Gap junction, Voltage-gated, second messenger-gated) help to sequester huge amounts of the ion in various organelles from where Ca^{2+} can be released upon extracellular stimuli. E.g. the contraction of the muscle is dependent on the presence of Ca^{2+} ions which are readily transported back into the organelles in order for the muscle to relax. In signal transduction, Ca^{2+} functions as a second messenger that activates Ca^{2+} dependent processes through the activation of Ca^{2+} /calmodulin dependent protein kinases (CaM kinases) which are the major effector molecules of Ca^{2+} . In the signaling cascades, the CaM dependent kinases activate phospholipases (e.g. phospholipase C) that in return activate other protein kinases such as protein kinase C.

Rab proteins

In eukaryotic cells the compartmentalization of processes is a prerequisite for a tight regulation of processes and activities. The cells contain a highly dynamic set of membrane compartments that are responsible for packaging, sorting, secreting, and recycling proteins and other molecules. Trafficking between organelles within the secretory pathway occurs as

vesicles derived from a donor compartment fuse with specific acceptor membranes, resulting in the directional transfer of cargo molecules. This process is tightly controlled by the Rab/Ypt family of proteins (reviewed by Novick and Zerial, 1997), a branch of the superfamily of small GTPases. Rab proteins regulate a variety of functions, including vesicle translocation and docking at specific fusion sites. Rabs may also play critical roles in higher order processes such as modulating the levels of neurotransmitter release in neurons, a likely mechanism in synaptic plasticity that underlies learning and memory (Geppert and Südhof, 1998).

Small GTPases share a common three-dimensional fold that, in the GTP bound state, can bind a variety of downstream effector proteins. GTP hydrolysis leads to a conformational change in the "switch" regions that renders the GTPase unrecognizable to its effectors. In this way, by localizing and activating a select set of effectors, a common structural motif is used to control a wide array of distinct cellular processes.

The final steps in membrane fusion are likely to be driven by a set of proteins known as SNAREs. After a vesicle becomes docked, the cytoplasmic domains of VAMP (also termed synaptobrevin) and syntaxin on opposing membranes, in combination with a SNAP-25 molecule, coalesce into an elongated -helical bundle (Poirier et al., 1998 ; Sutton et al., 1998), which may lead to fusion. Because numerous SNARE isoforms have been identified that localize to distinct membrane compartments, it was originally proposed that the specificity of interaction between the SNARE proteins accounted for the specificity in membrane trafficking. Recent results, however, suggest that SNAREs are not specific in their ability to form complexes in vitro, suggesting that trafficking specificity requires additional factors (Yang et al., 1999). In this regard, Rab proteins are strong candidates for governing the specificity of vesicle trafficking. Like the SNAREs, many isoforms (40) of the Rab family have been identified that localize to specific membrane compartments (reviewed by Novick and Zerial, 1997).

Concomitant with the SNARE cycle, Rab proteins undergo a intricate cycle of membrane and protein interactions. Rabs are posttranslationally modified at C-terminal cysteines by the addition of two geranylgeranyl groups, which mediate membrane association when the Rab is in the GTP-bound state. After guanine nucleotide hydrolysis occurs, the Rab is extracted from the membrane upon forming a complex with a cytosolic GDP-dissociation inhibitor (GDI). This cytosolic intermediate is then recycled onto a newly forming vesicle,

most likely through a secondary factor termed a GDI dissociation factor (GDF), which displaces GDI. After the Rab becomes membrane bound, a guanidine nucleotide exchange factor (GEF) promotes release of GDP and the subsequent loading of GTP. In its GTP-bound conformation, the Rab is then free to associate with its specific set of effectors, which can in turn trigger events leading to the eventual fusion of the vesicle with a target membrane. To complete the cycle, perhaps after or concurrent with membrane fusion, a GTPase activating protein (GAP) accelerates nucleotide hydrolysis, switching off the GTPase. The remaining GDP-bound Rab can then participate in a new round of fusion.

Rab interactions with effectors are likely to regulate vesicle targeting and membrane fusion in three ways. First, a Rab may specifically facilitate vectorial vesicle transport. Vesicles are transported from their site of origin to acceptor compartments likely through associations with cytoskeletal elements and transport motors. A protein has been identified with a domain structure that suggests a connection between the cytoskeleton and the Rabs. This protein, called Rabkinesin-6, contains a kinesin-like ATPase motor domain followed by a coiled-coil stalk region and a RBD that specifically binds Rab6 (Echard et al., 1998). An additional link with the cytoskeleton is provided by the Rab effector, Rabphilin-3A. Rabphilin-3A has been shown in vitro to interact with -actinin, an actin-bundling protein, but only when not bound to Rab3A (Kato et al., 1996). These results raise the intriguing possibility that Rab proteins regulate vesicle interactions with the cytoskeleton and thereby play an active role in targeting vesicles to their appropriate destinations.

Second, Rab proteins may regulate membrane trafficking at the vesicle docking step. A number of Rab effectors, including Rabaptin-5, EEA1, Rabphilin-3A, and Rim, may serve as molecular tethers. Each effector protein contains a RBD, followed by a linker region (some having the potential to form elongated coiled-coil structures), and a domain capable of interacting with a second Rab or the target membrane. Rabaptin-5, for example, contains two RBDs, one near the N terminus that specifically recognizes Rab4 and a second near the C terminus that binds Rab5 (Vitale et al., 1998). Both Rim, which is localized to the target membrane, and Rabphilin-3A, which is localized to the vesicle, contain N-terminal RBDs and C-terminal Ca^{2+} -binding C2 domains, implicating these effectors in synaptic vesicle localization or docking in response to Ca^{2+} influx (Wang et al., 1997). Tethering effectors may also recognize protein complexes on the acceptor membrane. Sec4p, a yeast Rab3A homolog, interacts with the exocyst (Guo et al., 1999), a complex of seven or more subunits

that is assembled at sites of vesicle fusion along the plasma membrane. The exocyst complex may therefore function as a landmark for Rab/effector-mediated vesicle docking.

Third, once a vesicle has become tethered to its fusion site, Rab proteins may selectively activate the SNARE fusion machinery. The mechanism of this activation is unknown but may involve direct interactions of Rabs or, more likely, their effectors with SNAREs. For example, Hrs-2 is a protein that binds to SNAP-25 and contains a Zn²⁺-finger motif characteristic of Rab-binding proteins such as Rabphilin-3A, Rim, EEA1, and Noc2, suggesting that Hrs-2 may form a physical link between Rabs and SNAREs (Bean et al., 1997). In addition, certain mutations in the syntaxin-binding protein Sly1p, the Sec1p homolog utilized in ER to Golgi trafficking, eliminate the requirement for Ypt1p, a Rab protein that functions at this trafficking step (Dascher et al., 1991). Rabs may therefore regulate SNARE associations through Sec1 family members. In support of this idea, a Rab effector was recently found to interact with a vacuole Rab, a Sec1p homolog, and a SNARE protein (Peterson et al., 1999), which suggests that this effector serves to connect Rab and SNARE function. In this way, Rabs and their effectors may facilitate the correct pairing of SNAREs.

References: Dascher et al. (1991). *Mol. Cell. Biol.* 11, 872-885; Echard et al. (1998). *Science*. 279, 580-585; Geppert et al. (1998). *Annu. Rev. Neurosci.* 21, 75-95; Guo et al. (1999). *EMBO J.* 18, 1071-1080; Kato et al. (1996). *J. Biol. Chem.* 271, 31775-31778; Novick et al. (1997). *Curr. Opin. Cell Biol.* 9, 496-504; Peterson et al. (1999). *Curr. Biol.* 9, 159-162; Poirier et al. (1998). *Nat. Struct. Biol.* 5, 765-769; Vitale et al. (1998). *EMBO J.* 17, 1941-1951; Wang et al. (1997). *Nature*. 388, 593-598; Yang et al. (1999). *J. Biol. Chem.* 274, 5649-5653.

Kinases

Reversible posttranslational modifications of proteins are major means of regulating cellular activities. Among the various modifications that are carried out by the cells, the addition of phosphoryl groups to Ser/Thr or Tyr residues is the most important and widely used. The phosphorylation of proteins is accomplished by protein kinases, while the reverse reaction, the removal of phosphoryl groups, is carried out by phosphatases. Kinases / Phosphatases regulate key positions e.g. in the processes of cell proliferation, differentiation and communication/signaling. These processes must be tightly regulated in order to maintain a steady state level of cellular fate. Mis-regulation of kinase activities (or that of

phosphatases) is made responsible for a multitude of disease processes such as oncogenesis, inflammatory processes, arteriosclerosis, and psoriasis.

Protein kinases constitute the largest protein family that is currently known. Several hundred kinases have been identified already. Classically, kinases are subdivided into two classes based on the amino acid residues in their substrates that are phosphorylated by the particular enzymes. The kinases specifically add phosphoryl groups from adenosine triphosphate (ATP) or, less frequently, guanosine triphosphate (GTP), either to serine and/or threonine or to tyrosine residues of substrate proteins. An estimated 1,000 to 10,000 proteins present in a typical mammalian cell are believed to be regulated also by the action of protein kinases.

Protein kinases are frequently integral parts of signaling cascades that transmit extracellular stimuli (e.g. hormones, neurotransmitters, growth- or differentiation factors) into the cell and result in various responses by the cells. The kinases play key roles in these cascades as they constitute a sort of 'molecular switches' turning on or off the activities of other enzymes and proteins, e.g. metabolic, regulatory, channels and pumps, receptors, cytoskeletal, transcription factors.

The regulation of kinase activities is accomplished by various means:

The best characterized example for the regulation via regulatory subunits is the cAMP-dependent protein kinase (PKA) which is also a prototype for second messenger activated protein kinases. This enzyme consists of a heterotetramer of two catalytic (C) and two regulatory (R) subunits. Upon binding of two molecules of second messenger (cAMP) in each R subunit, the catalytic subunits are released and active. Both of the catalytic and the regulatory subunits several isoforms exist. The combination of catalytic and regulatory subunits determines the localization of the holoenzyme and also the substrate spectrum that is available for phosphorylation. The consensus pattern necessary to be present in the substrate for PKA action is RRXS/T where X can be any amino acid.

The casein kinase II comprises another examples for holoenzymes that consist of catalytic and regulatory subunits. Other kinases that are activated by second messengers are cGMP-dependent protein kinase and Protein kinase C (PKC) which is activated by diacylglycerol, which in turn is produced by phospholipases by cleavage of phosphatidylcholine.

Receptor kinases usually consists of an extracellular domain which can bind effector molecules (e.g. growth factors and hormones) and transfer the stimulus to the intracellular domain of these proteins which usually is a protein tyrosine kinase. Other tyrosine kinases lack an extracellular domain but are associated with receptors which transfer the signal after effector binding by activating the associated protein kinase enzyme (e.g. Src kinase family; Src, Blk, Fgr, Fyn, Lck Lyn, Yes and Janus kinase family; Jak1-3, Tyk2).

Dysfunction of kinases, e.g. caused by non-functioning regulation, can be the cause of inflammatory diseases and uncontrolled proliferation. v-Src which is a truncated version of the C-Src protooncogene tyrosine kinase is a classical example for this process as v-Src does not contain the regulatory domain of the cellular gene and is thus constitutively active.

Several categories of proteins are coded for by clones of the invention within the overall group of "Signal transduction" and include, among others, the following:

Neurocalcin (Recoverin): Neurocalcin is a Ca^{2+} -binding protein with three putative Ca^{2+} -binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca^{2+} dependent activation of guanylate cyclase. These proteins can find application in modulating/blocking the guanylate cyclase-pathway. Diseases associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) autosomal dominant cone dystrophy (OMIN *600364); 2) cone dystrophy 3 (OMIN *600364); 3) cancer associated retinopathy (OMIN *179618). Clones in this category include: fbr2_23b21.

Proteins with a WW Domain: Proteins that contain a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore these proteins should be involved in intracellular signal transduction. Diseases associated (as

potentially diagnostic, therapeutic, causative, and/or related, etc...) with these proteins include as reported by OMIN 1) Muscular Dystrophy, Pseudohypertrophic Progressive Duchenne and Becker Types (OMIN *310200). Clones in this category include: fbr2_23n16.

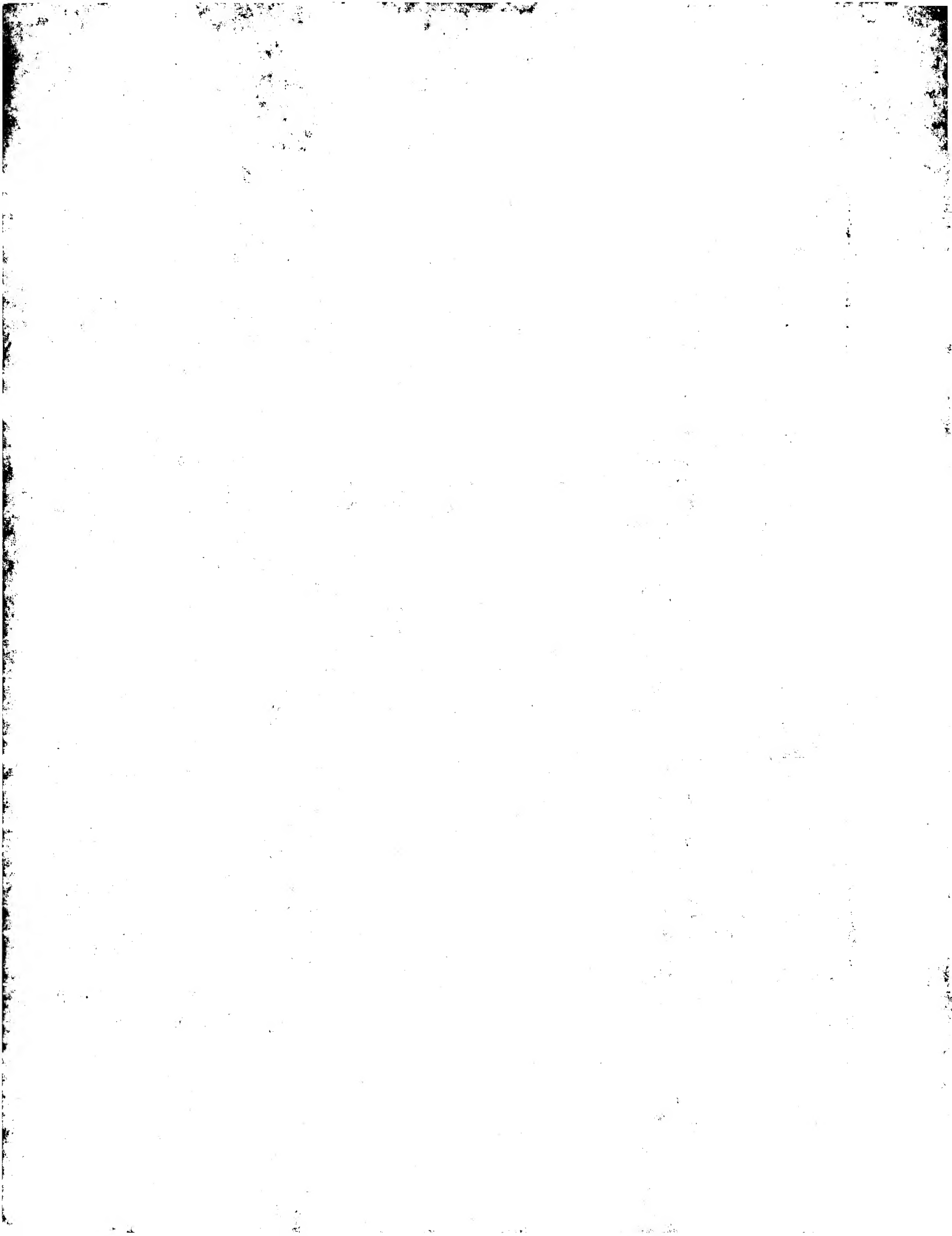
Protein substrates for cAMP-dependent protein kinase: Acting as a choride channel or chloride channel inhibitor these proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Cystic Fibrosis (OMIN #219700). Clones in this category include fbr2_82i17.

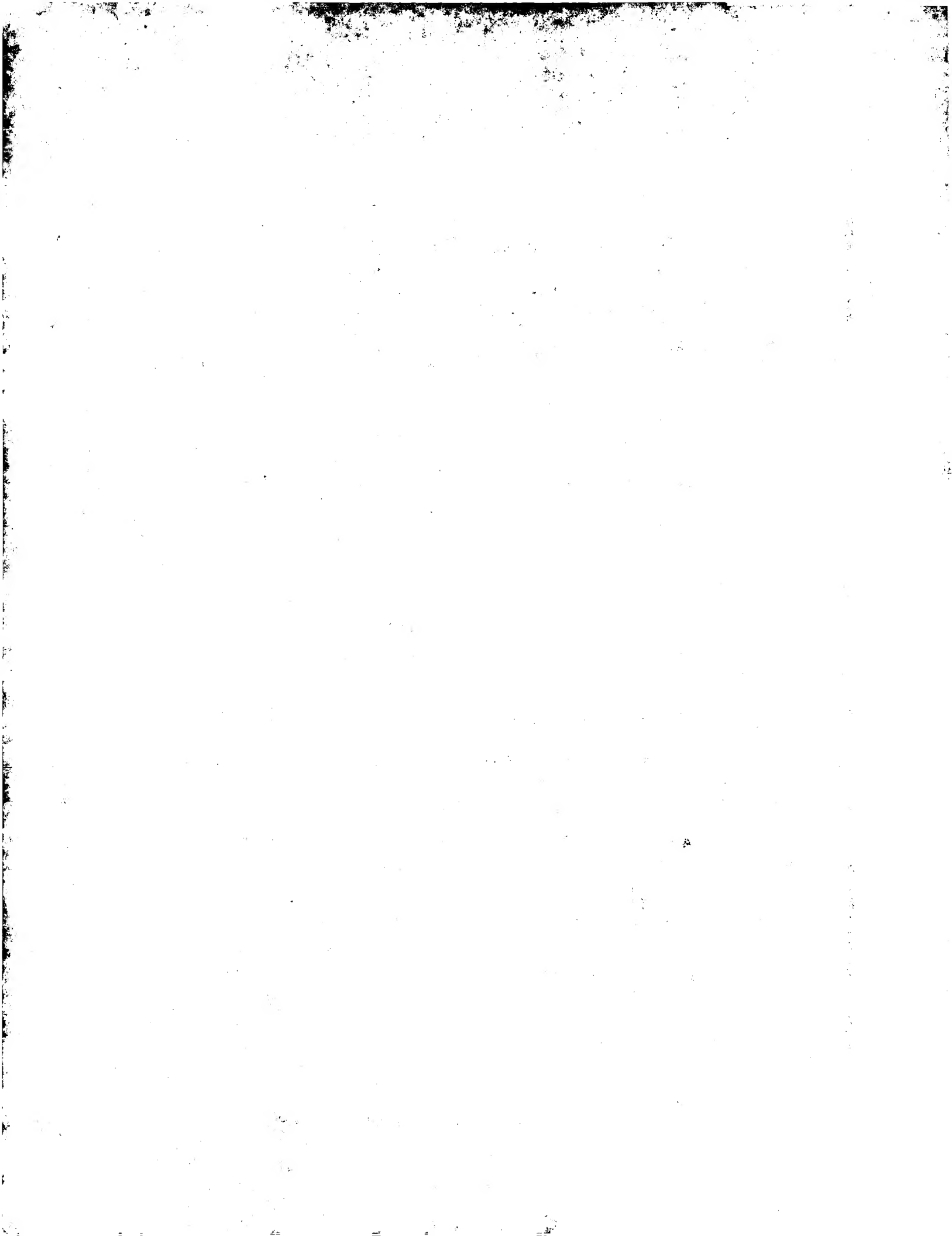
Sphingosine kinase: Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellular, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependent on SPP. Extracellularly, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1. These proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) as reported by OMIN with Gaucher Disease, Type I (OMIN *230800). Clones in this category include fbr2_82m6.

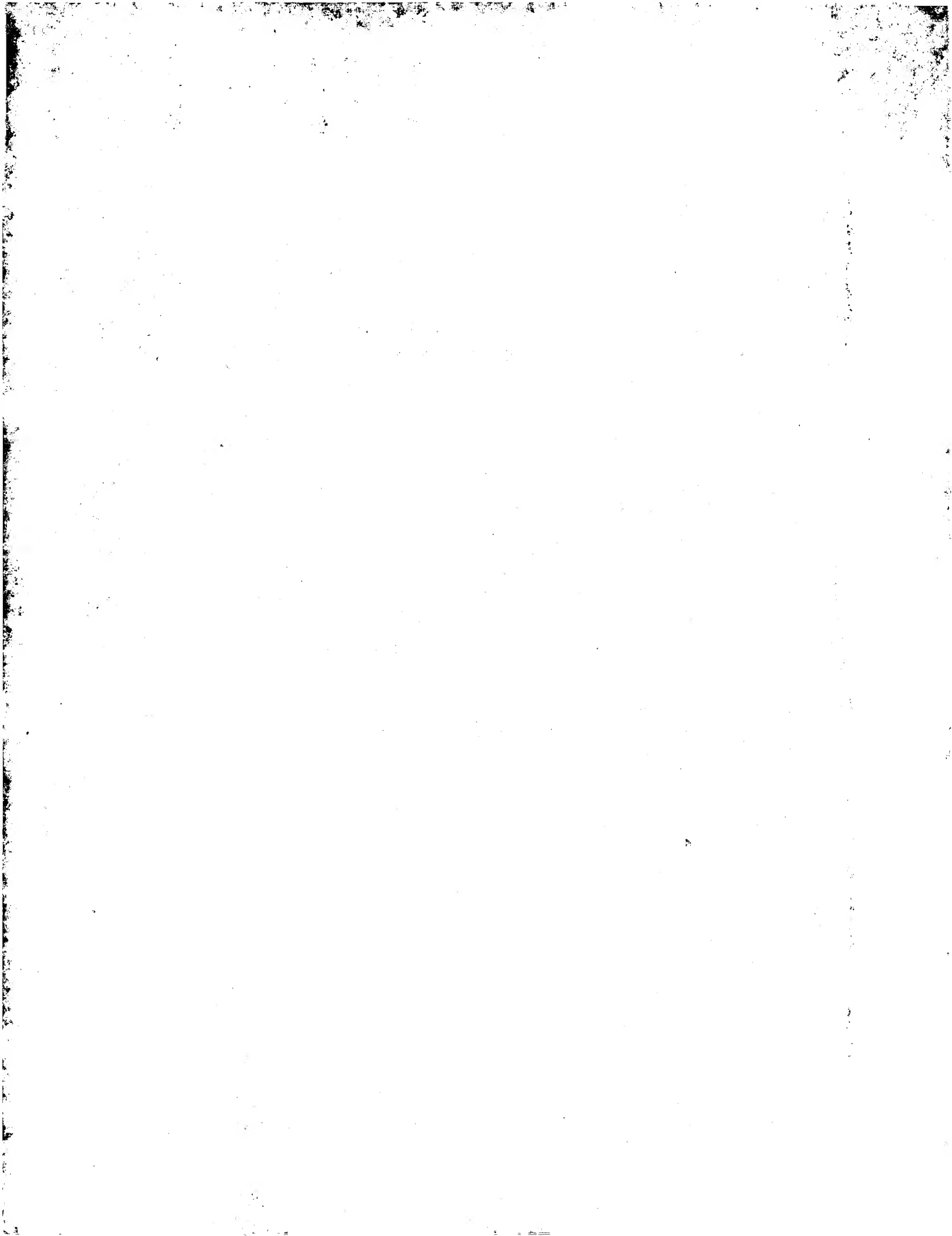
Vanilloid Receptors: VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. Related can find application as a target for the development of new nociception-modulating drugs. Clones in this category include tes3_20k2.

RCC1 (Regulator of chromosome condensation): RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator. These proteins can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat. OMIN also reports that RCC1 has associations (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with retinitis pigmentosa (OMIN *312610). Clones in this category include tes3_21d4.

Ras inhibitor proteins: Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show







intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. Ras inhibitor proteins have been associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with many disease processes as reported by OMIN including: 1) Tumors of the lung, breast, brain, pituitary, pancreas, bone, skin, bladder, kidney, ovary, prostate and lymphocyte, Melanoma (OMIN *600160); 2) X-linked non-specific mental retardation (OMIN *300104); 3) adenomatous polyposis of the colon (OMIN *175100); 4) Beckwith-Wiedemann Syndrome (#130650); and 5) Major affective disorder 1 (OMIN *125480). Clones in this category include ute1_22g21.

Mammalian proteins cornicon involving the EGF-receptor: Cornicon proteins are part of a signal transduction pathway involving the EGF-receptor. The EGF-receptor has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) Familial hypercholesterolemia (OMIN 143890); 2) Leprechaunism (OMIN #246200); 3) Hemophilia B (OMIN *306900); 4) Ectodermal dysplasia 1; 5) Kartagener syndrome (OMIN *244400) and 6) Glioma of the brain (OMIN *137800). Clones in this category include ute1_22e12.

Transmembrane proteins

Membrane region prediction was effected using the ALOM2 software (Klein et al., 1985; version 2 by K. Nakai). Similar to many other methods, the Kyte & Doolittle (1982) amino acid hydrophobicity scale is used in ALOM2 as the primary variable for classifying sequences in terms of their localization. High prediction accuracy is achieved through the system of intelligent decision rules and the utilization of a carefully selected training data set. The method also generates reliability estimates which makes it possible to distinguish between membrane-spanning proteins (I, intrinsic) and globular proteins with regions of high hydrophobicity buried in the core.

For a protein of length L , the block of length l with maximum hydrophobicity is found:

$$\max H = \max_{k=1, \dots, L-l+1} (1/l) \sum_{i=k}^{k+l-1} H_i$$

where H_i represents the hydrophobicity of an individual residue.

Let $P(I/\max H)$ and $P(E/\max H)$ be the conditional probabilities that a protein is integral or peripheral, respectively, given its value of maximal hydrophobicity $\max H$, and let $P(I)$ and $P(E)$ be the prior probabilities of intrinsic and extrinsic membrane proteins estimated from the training set. Then a sequence is assigned to E if

$$P(E/\max H) > P(I/\max H)$$

or, after applying the Bayes rule,

$$P(E)P(\max H/E) > P(I)P(\max H/I),$$

where the conditional probabilities $P(\max H/E)$ and $P(\max H/I)$ can be determined based on the estimates of probability distributions of $\max H$ in both groups.

Discriminant analysis allows to simplify this task by calculating the odds $P(E/\max H):P(I/\max H)$ as e^b , where b is the left-hand side of a linear or quadratic inequality. For example, for the window of length 17, the protein is allocated to the peripheral category E based on the empirically derived quadratic inequality:

$$1.05(\max H)^2 + 12.30\max H + 17.49 > 0,$$

whereas the optimal inequality for assigning membrane proteins (category I) is linear:

$$-9.02\max H + 14.27 > 0$$

The odds parameter can be made more or less stringent. For example, one can require odds at least 1:10 for a protein to be classified as integral. This leads to higher selectivity but less sensitivity.

The boundaries of membrane-spanning regions in putative membrane proteins are detected by means of an iterative procedure whereby the most hydrophobic region corresponding to the value $\max H$ is considered to be membrane and removed from the sequence. The classification procedure is then repeated again for the remaining sequence, and, if such a protein is again classified as integral, the next most hydrophobic region is considered.

Reference: Klein, P., Kanehisa, M., DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochem Biophys Acta* **815**: 468-476

Transcription factors

Purified eukaryotic RNA polymerase II is unable to initiate promoter-specific transcription. A family of factors that collectively confer RNAPII promoter specificity is known as the general transcription factors (GTFs). They include the TATA-binding Protein (TBP) TFIIB, TFIIE, TFIIIF and TFI IH. These factors are conserved among all eukaryotes.

RNAPII complexes containing the entire set of GTFs or a subset of GTFs together with other proteins have been isolated from mammalian and yeast cells. Although purified RNAPII and GTFs are sufficient for promoter-specific initiation, this system fails to respond to activators. This is mediated by a further complex termed mediator complex which associates with the carboxy-terminal heptapeptide domain (CTD) of the largest subunit of RNAPII.

Purification of human RNAPII complexes resulted in two distinct forms of human RNAPII after analysis of functional properties. One complex contained chromatin remodeling activities but was devoid of GTFs. The other complex did not contain factors that modify chromatin but contained a subset of SRB/mediator subunits and GTFs and other polypeptides that mediate transcriptional activation, a scenario similar to that reported for yeast.

A complex designated NAT (~20 SU) for negative regulator of transcription contains RNAPII, Cdk8, homologs of the yeast mediator complex as well as Rgr1 and Srb10/11 known as negative regulators of transcription.

A complex with striking similar structural and functional properties to NAT has been identified designated SMCC (~15 SU) (SRB/mediator coactivator complex), that can also mediate transcriptional activation.

The SMCC complex includes all reported NAT subunits including subunits of the TRAP complex. TRAP is a coactivator complex isolated on the basis of its interaction with the thyroid hormone receptor. Another coactivator complex DRIP, isolated on the basis of its

ability to interact with the vitamin D3 receptor, contains novel subunits as well as subunits of NAT/SMCC and TRAP complexes.

The effects of each of these coactivator complexes is dependent on the TFIID complex. It is not known if the TAF subunits of TFIID are required. It is likely that new coactivator complexes will be uncovered containing both novel and previously defined components.

Beside the huge amount of transcription factors which can be part of the RNAIIP holoenzyme or the coactivator complexes there is an even larger quantity of specific transcription factors binding to promoter elements within the DNA sequences of a given gene leading to activation or repression of transcription. A broad range of cellular responses like differentiation, proliferation, cell death and others are elicited through activating or repressing the transcription of target genes.

There are at least five superclasses of transcription factors:

1. Superclass contains members with characteristic basic domains:

Members are:

Leucine zipper factors, where the basic domain is followed by a leucine zipper of repeated leucine residues at every seventh position. The zipper mediates protein dimerization as a prerequisite for DNA-binding.

Helix-loop-helix factors (bHLH) contain a DNA-binding basic region followed by a motif of two potential amphipathic alpha-helices connected by a loop of variable length also mediating dimerization.

Factors with a combination of Helix-loop-helix and leucine zipper.

Further members of this superclass are NF- κ B, RF-X, and bHLH like proteins.

2. Superclass comprises factors containing zinc-coordinating DNA-binding domains.

Members are:

Proteins with Cys4 zinc finger of nuclear receptor type, where two such motifs differing in size, composition and function are present in each receptor molecule. Each finger comprises 4 cysteine residues coordinating one zinc ion. The second half including the second cysteine pair has alpha-helix conformation and the helix of the first finger binds to the DNA through the major groove. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding. This class includes the steroid hormone receptors and the thyroid hormone receptor-like factors. Other diverse cys4 zinc fingers have a motif of GATA-type.

Proteins with Cys2His2 zinc finger domain(s). Each finger comprises 2 cysteine and 2 histidine residues coordinating one zinc ion, and in some cases one histidine is replaced by another cysteine. The zinc ion is essential for DNA-binding.

Proteins with Cys6 cysteine-zinc cluster(s). Six cysteine residues coordinate two zinc ions, i. e. two of the thiol groups are coordinating two zinc ions each. Present in many fungal regulators.

Zinc fingers of alternating composition.

3. Superclass contains factors of helix-turn-helix type.

Members are:

Proteins with homeo domains. Homeo domains are three consecutive alpha-helix structures. Helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.

Proteins with Paired box domain(s). This is a DNA-binding domain of approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half is highly charged in general. It probably comprises 3 alpha-helices.

Proteins with Fork head / winged helix domain(s). This domain was identified by homology between HNF-3A and fkh. The domain comprises approx. 110 AA. Analysis of the crystal structure has revealed a compact structure of three alpha-helices, the third alpha-helix

being exposed towards the major groove of the DNA. The domain also exerts minor groove contacts. Upon binding to DNA, it induces a bend of 13 degree.

Heat shock factors

Proteins with Tryptophan clusters. The tryptophan clusters comprise several tryptophan residues with a spacing of 12-21 amino acid residues; the subclass of myb-type DNA-binding domains typically exhibit a spacing of 19-21 amino acid residues.

Proteins with TEA domain(s). The TEA domain has been identified as a region which is conserved among the transcription factors TEF-I, TECl and abaA. This domain in TEF-I has been shown to interact with DNA, although two additional regions may also contribute to DNA-binding. It is predicted to fold into three alpha-helices, with a randomly coiled region of 16-18 amino acid residues between helices 1 and 2, and a short stretch between helices 2 and 3 of 3-8 residues.

4. Superclass contains beta-Scaffold Factors with Minor Groove Contacts

Members are:

Proteins with RHR (Rel homology) region.

The structure of the Rel-type DBD exhibits a bipartite subdomain structure, each subdomain comprising a beta-barrel with five loops that form an extensive contact surface to the major groove of the DNA. Particularly, the first loop of the N-terminal subdomain (the highly conserved recognition loop) performs contacts with the recognition element on the DNA, but other loops are involved. The fact that the main DNA-contacts are made through loops has been suggested to provide a high degree of flexibility in binding to a range of different target sequences. Augmenting interactions are achieved by two alpha-helices within the N-terminal Part that form strong minor groove contacts to the A/T-rich center of the B-element. In p65, the sequence between both alpha-helices is much shorter and even helix 2 is truncated. The second, C-terminal domain is necessary mainly for protein dimerization.

p53 proteins

MADS (MCM1-agamous-deficiens-SRF) box proteins. Proteins of this class comprise a region of homology. The DNA-binding domain also comprises the dimerization capability. In the DNA-bound dimer (shown for SRF), two antiparallel amphipathic alpha-helices (alpha-I), form a coiled coil and are oriented approximately parallel on the minor groove. These helices make minor and major groove contacts, the N-terminal extensions form minor groove contacts. The bound DNA is bent and wrapped around the protein. It exhibits a compressed minor groove in the center and widened minor groove in the flanks.

Beta-Barrel alpha-helix transcription factors.

TATA-binding proteins

HMG proteins

Proteins of this class comprise a region of homology with the chromosomal non-histone HMG proteins such as HMG1. This region comprises the DNA-binding domain which in some instances such as HMG1 mediates sequence-unspecific, in other cases such as LEF-1 sequence-specific binding to DNA. This domain exhibits a typical L-shaped conformation made up of 3 alpha-helices and an extended N-terminal extension of the first helix. The latter together with helix 1, which contains a kink, form the long arm of the L, whereas helices 1 and 2 form the short arm. Binding to the minor groove induces a sharp bending of the DNA by more than 90 degree, away from the bound protein. The overall topology of the DNA-protein complexes resembles somewhat that of the TBP-TATA box complex.

Heteromeric CCAAT factors

Proteins with Grainyhead domain(s)

Cold-shock domain factors. Cold-shock domain proteins are characterized by a highly conserved region first found in prokaryotic cold-shock proteins. This domain is a single-stranded nucleic acid-binding structure interacting with DNA or RNA. It consists of an antiparallel five-stranded beta-barrel, the strands of which are connected by turns and loops. Within this structure, a three-stranded beta-strand contains a conserved RNA-binding motif, RNPI. Not all CSD proteins are transcription factors. Those which specifically bind to a

certain sequence are termed Y-box proteins. Proteins of this class were previously called protamine-like domain proteins because of having a highly positively charged domain with interspersed proline residues.

Proteins with Runt homology domain

The members of this transcription factor class have been identified on the basis of their homology to a defined region within the *Drosophila* protein Runt. The runt domain is part of the DNA-binding domain of these factors. It consists mainly of beta-strands, does not contain alpha-helical regions and seems to be most similar to the palm domain found in DNA polymerase beta (rat).

5. Superclass contains other transcription factors like Copper fist proteins, HMGI(Y), STAT, Pocket domain proteins and Ap2/EREBP-related factors.

The classification of transcription factors originates from TRANSFAC database:

<http://transfac.gbf.de/TRANSFAC/>

Reference: Heinemeyer

Several categories of proteins are coded for by clones of the invention within the overall group of "Transcription Factors".and include, among others, the following:

Dcoh: Dcoh is a bifunctional protein, complexed with bipterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bipterin cofactor of phenylalanine hydroxylase. The Dcoh protein has been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) hyperphenylalanemia (OMIN 126090, #264070). Clones in this category include fkd2_46k12.

Signal transducing proteins: Beta-transducin subunits of G-proteins contain WD-40 repeats. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription. These proteins have been reported by OMIN to be associated (as potentially diagnostic, therapeutic, causative, and/or related, etc...) with the following diseases: 1) essential hypertension (OMIN *139130). Clones in this category include ute1_1i2.

* * *

The invention, therefore, specifically contemplates the following assemblages of materials, which track the above-identified fourteen functional groupings, that are useful in practicing the profiling aspects of the invention. One type of assemblage is nucleic acid-based and can include the following groupings of sequences and their derivatives: all sequences; human fetal brain sequences; brain derived sequences; human fetal kidney library sequences; kidney derived sequences; human mammary carcinoma library sequences; mammary carcinoma derived sequences; human testis library sequences; testes derived sequences; cell cycle genes; cell structure and motility genes; differentiation and development genes; intracellular transport and trafficking genes; metabolism genes; nucleic acid management genes; signal transduction genes; transmembrane protein genes; and transcription factor genes. Other assemblages contain proteins or their corresponding antibodies or antibody fragments, divided along the same groupings.

Database Applications

Because they are human genes and gene products, the inventive molecules are useful as members of a database. Such a database may be used, for example, in drug discovery and rationale drug design or in testing the novelty and non-obviousness of newly sequenced materials. In addition, they are particularly suited in designing variants for the profiling (and other) applications described herein. Hence, the following discussion of electronic embodiments applies equally to such variants, which, naturally, will be generated and stored using a computer using known methodologies.

Accordingly, one aspect of the invention contemplates a database of at least one of the inventive sequences stored on computer readable media. Again, the individual sequences may be grouped with regard to the individual functional and structural groups mentioned above. While the individual sequences of a database may exist in printed form, they are preferably in electronic form, as in an ascii or a text file. They may also exist as word processing files or they may be stored in database applications like DB2, Sybase, Oracle, GCG and GenBank. One skilled in the art will understand the range of applications suitable for using and storing the electronic embodiments of the invention.

"Computer readable media" refers to any medium which can be read and accessed by a computer. These include: magnetic storage media, like floppy discs, hard drives and magnetic tape; optical storage media, like CD-ROM; electrical storage media, like RAM

and ROM; and hybrids of these categories, like magnetic/optical storage media. One skilled in the art will readily understand the scope of computer readable media and how to implement them.

Biological Activities and Assays for Implementing Therapeutic and Diagnostic Applications

This section provides assays for biological activity that are useful in characterizing and quantifying the biological activity of the inventive molecules and their derivatives, which is relevant to the pharmacological effects of the inventive molecules. As used in this section, it will be understood that "protein" may also refer to the inventive antibodies (including fragments).

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M + (preB M +), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin gamma, Schreiber, R. D. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., *J. Exp. Med.* 173:1205-1211, 1991; Moreau et al., *Nature* 336:690-692, 1988; Greenberger et al., *Proc. Natl. Acad. Sci. U.S.A.* 80:2931-2938, 1983; Measurement of mouse and human interleukin 6-Nordan, R. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., *Proc. Natl. Acad. Sci. U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger et al., *Eur. J. Immun.* 11:405-411, 1981; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to modify immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the

tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this manner prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient.

The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and beta 2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Herrmann et al., *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann et al., *J. Immunol.* 128:1968-1974, 1982; Handa et al., *J. Immunol.* 135:1564-1572, 1985; Takai et al., *J. Immunol.* 137:3494-3500, 1986; Bowman et al., *J. Virology* 61:1992-1998; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnoli et al., *Cellular Immunology* 133:327-341, 1991; Brown et al., *J. Immunol.* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; and Assays for B cell function: *In vitro* antibody production, Mond, J. J. and Brunswick, M. In *Current Protocols in Immunology*. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai et al., *J. Immunol.* 137:3494-3500, 1986; Takai et al., *J. Immunol.* 140:508-512, 1988; Bertagnoli et al., *J. Immunol.* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., *J. Immunol.* 134:536-544, 1995; Inaba et al., *Journal of*

Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad. Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the

treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and

cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. *Invest. Dermatol* 71:382-84 (1978).

Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle

stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin alpha family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- beta group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of

cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such

receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltzenberg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of

cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in

a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Particular Applications for Certain Clones

The following sets out a non-exclusive list of applications for certain embodiments of the invention. In the interest of economy, applications relevant to multiple embodiments are not duplicated in this list. Other embodiments described in below have similar characteristics, as described therein. The artisan is directed, therefore, to this section for similar descriptions of the functions of other embodiment.

Testes

htes3_15c24: The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

htes3_15i5: The new protein can find application in modulating the structure of the human spermatozoa radia spoke head and modulation of sperm motility in men.

htes3_15k11: The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

htes3_17n12: The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

htes3_20k2: The new protein can find application as a target for the development of new nociception-modulating drugs.

htes3_20m18: The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

htes3_20d4: The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

htes3_21j15: NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor. The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

htes3_26g22: The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP. The new protein can find application in modulation of gene transcription.

htes3_21l16: The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

htes3_27d1: The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

htes3_2m18: The novel protein can find application as multifunctional nuclease / exoribonuclease.

htes3_35b4: The new protein can find application in modulation of the mitotic spindle.

htes3_35b5: The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

htes3_35e21: Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells. This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

htes3_35k16: Therefore it is a new fatty acid-CoA synthetase/ligase with unknown substrate. The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

htes3_35n12: The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

htes3_35n9: The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

htes3_35p22: The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control. The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

htes3_4h6: The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

htes3_72k15: FGD1-related F-actin-binding protein (Frabin/FGD1) is a novel F-actin-binding protein. The gene locus *fgd1* seems to be responsible for faciogenital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GDP/GTP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an *esin* yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

htes3_72p16: As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (PrA), proteinase B (PrB), and alkaline phosphatase (ALP). The new protein can find application in modulation the sorting of proteins into different compartments.

htes3_7b22: The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

htes3_7j3: The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too. The new protein can find application in modulating/blocking the cell cycle.

htes3_7p9: The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle. The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819) complex. The new protein can find application in modulation of viral infections and tumour events.

htes3_8m10: The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA. The new protein can find application in modulation of mRNA translation and processing/stability.

Kidney

hfk2_24b15: The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

hfk2_24n20: The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton. The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

hfk2_3o17: The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

hfk2_46j20: The new protein can find application in modulating the homoprotocatechuate degradative pathway and as an enzyme for biotechnologic production processes.

hfk2_46k19: The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

hfk2_46m4: SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

hfk2_46k14: rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport. The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

Uterus Associated:

hutel_18i19: The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

hutel_18i1: The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome sub-unit.

hutel_19g22: The new protein can find application in modulation of tissue-calcification, especially the uterus.

hutel_19h17: The new protein can find application in modulating the response of cells to oxysterols.

hutel_20b19: The novel protein seems to be a novel enzyme with sarcosine oxidase activity. The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

hutel_20g21: The novel protein seems to be a new ras inhibitor protein. The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

hutel_20h13: The novel protein is a new human alpha-adaptin. The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

hutel_20m11: The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

hutel_20m24: This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2. The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

hutel_22e12: The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

hutel_23e13: The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnologic production processes.

hutel_24j6: The new protein can find application in modulation of cell-cell-adhesion.

hutel_24h3: The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

Fetal Brain:

hfbr2_16c16: The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

hfr2_23b21: The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

hfr2_23b10: The new protein can find application in modulation of splicing.

hfr2_2b5: The novel protein contains the typical (xxG)_n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain. The new protein can find application in modulation of connective tissue, bone and cartilage development and maintenance.

hfr2_2c17: The new protein can find application in modulating/blocking G-protein-dependent pathways.

hfr2_2d15: The new protein can find application in modulating early spermatogenesis.

hfr2_2i17: The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

hfr2_2k14: Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumour-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

hfr2_3c18: RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

hfr2_3g8: The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

hfbr2_62b11: The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

hfbr2_62o17: The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins.

hfbr_6b24: The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

hfbr_72b18: The new protein can find application in modulating DNA repair and mutagenesis.

hfbr_78c4: The new protein can find application in modulating/blocking the response of cells to interferons.

hfbr_78k24: These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. The new protein can find application in modulation of protein stability/degradation in cells.

hfbr_82e4: The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

VARIANTS OF THE INVENTIVE DNA MOLECULES

Variants in General

"Variants," according to the invention, include DNA and/or protein molecules that resemble, structurally and/or functionally, those set forth in herein. Variants may be isolated from natural sources ("homologs"), may be entirely synthetic or may be based in part on both natural and synthetic approaches.

The section set forth below presents various structural and functional characteristics of molecules within the invention. Preferred molecules are characterized by a combination of one or more of these characteristics. For instance, some preferred molecules are described with reference to at least two structural characteristics, while others may be described with reference to at least one structural and at least one functional characteristic.

It will be recognized by the skilled artisan that structure ultimately defines function, *i.e.* the functions of the molecules described herein derives from the structures of those

molecules. Accordingly, the structural variants described below that bear the closest structural relationship (as variously defined below) to the inventive molecules are the variants that most likely will preserve biological function. This relationship between structure and function will guide the skilled artisan in identifying the preferred embodiments of the invention.

Splicing Variants

It is well-known that eukaryotic structural genes are comprised of both protein coding and non-coding portions. When the messenger RNA is transcribed from the DNA template, it contains introns, which are non-coding, and exons, which are coding. In order to form a translation competent mRNA, the introns must be "spliced" out of this initial pre mRNA.

Specific sequences within the pre mRNA represent "splice junctions" that direct the cellular splicing machinery to the appropriate position. The splice junctions are loosely conserved sequence regions of the pre mRNA, which almost invariably begin with GT and end with AG (DNA perspective). The 5' end of the splice junction typically contains about nine somewhat conserved residues, for example, C/AAGTA/GAGT. The 3' end usually contains a pyrimidine rich stretch of at least about 11 nucleotides, followed by NC/TAGG. Splicing occurs before the GT and after the AG. Mount, *Nucleic Acids Res.* 10:459-72 (1982).

Interestingly, exons often correspond to discrete functional domains of the protein product. The intron/exon arrangement thus creates a linear array of nucleotides which can be correlated to discrete, and often interchangeable, functional protein fragments. Go, *Nature* 291:90-92 (1981); Branden *et al.*, *EMBO J.* 3:1307-10 (1984). This linear arrangement creates the possibility of generating multiple different full length proteins by rearranging the order of the different functional portions in the array. For example, if a set of exons are arranged 1-2-3-4, where (-) represents the introns separating the exons, a splicing event need not simply produce 1234, but may produce 123, 134, 124 and so on. Production of different mRNA products in this way is commonly called "alternative splicing." Andreadis *et al.*, *Ann. Rev. Cell Biol.* 3:207-42 (1987).

Some of the present DNA molecules can be represented in modular fashion in terms of their coding regions. Essentially, these modules are exons (though each "exon" may in fact be made up of several exons), which may be combined in different ways to form a variety of

different DNA molecules, each encoding a different functional protein. Splicing variants are indicated below.

Degenerate Variants

One aspect of the present invention provides "degenerate variants" of the nucleic acid fragments of the present invention. A "degenerate variant" is a nucleotide fragment which differs from those of inventive molecules by nucleotide sequence, but due to the degeneracy of the genetic code, encodes an identical polypeptide sequence.

Given the known relationship between DNA sequences and the proteins they encode, degenerate variants typically are described by reference to this relationship. It is well known that the degeneracy of the genetic code results in many possible DNA sequences which encode a particular protein. Indeed, of the three bases which comprise an amino acid-encoding triplet, the third position, and often the second, almost always may vary. This fact alone allows for a class of variant DNA molecules which encode protein sequences identical to those disclosed herein, yet have about 30% sequence variation. In other words, the variant DNA molecules are about 70% identical to the inventive DNAs, having no additional or deleted sequences. Thus, one aspect of the invention provides degenerate variant DNA molecules encoding the inventive protein sequences.

In one embodiment, these variants have at least about 70% sequence identity with the DNA molecules described herein. In a preferred embodiment, these variants have at least about 80% sequence identity to the inventive molecules. In a more preferred embodiment these variants have at least about 90% sequence identity with the inventive molecules.

Conservative Amino Acid Variants

Variants according to the invention also may be made that conserve the overall molecular structure of the encoded proteins. Given the properties of the individual amino acids comprising the disclosed protein products, some rational substitutions will be recognized by the skilled worker. Amino acid substitutions, *i.e.* "conservative substitutions," may be made, for instance, on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example: (a) nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; (b) polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine;

(c) positively charged (basic) amino acids include arginine, lysine, and histidine; and (d) negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Substitutions typically may be made within groups (a)-(d). In addition, glycine and proline may be substituted for one another based on their ability to disrupt α -helices. Similarly, certain amino acids, such as alanine, cysteine, leucine, methionine, glutamic acid, glutamine, histidine and lysine are more commonly found in α -helices, while valine, isoleucine, phenylalanine, tyrosine, tryptophan and threonine are more commonly found in β -pleated sheets. Glycine, serine, aspartic acid, asparagine, and proline are commonly found in turns. Some preferred substitutions may be made among the following groups: (i) S and T; (ii) P and G; and (iii) A, V, L and I. Given the known genetic code, and recombinant and synthetic DNA techniques, the skilled scientist readily can construct DNAs encoding the conservative amino acid variants.

As used herein, "sequence identity" between two polypeptide sequences indicates the percentage of amino acids that are identical between the sequences. "Sequence similarity" indicates the percentage of amino acids that either are identical or that represent conservative amino acid substitutions.

Functionally Equivalent Variants

Yet another class of DNA variants within the scope of the invention may be described with reference to the product they encode. As shown below, some of the inventive DNA molecules encode a protein having a degree of homology with known proteins, or protein domains. It is expected, therefore, that they will have some or all of the requisite functional features of such molecules. These "functionally equivalent variants" products are characterized by the fact that they are functionally equivalent, with respect to biological activity, to certain known molecules.

The instant invention provides information on common structural motifs, including consensus sequences that will guide the artisan in constructing functionally equivalent variants. It will be understood that the motifs, identified for each inventive protein, may be modified within the identified consensus sequences. Thus, the invention contemplates the proteins disclosed herein that contain variability in the consensus sequences identified, and the invention further contemplates the full range of nucleic acids encoding them, and the complements of those nucleic acids.

Hybridizing Variants

DNA variants within the invention also may be described by reference to their physical properties in hybridization. One skilled in the field will recognize that DNA can be used to identify its complement and, since DNA is double stranded, its equivalent or homolog, using nucleic acid hybridization techniques. It will also be recognized that hybridization can occur with less than 100% complementarity. However, given appropriate choice of conditions, hybridization techniques can be used to differentiate among DNA sequences based on their structural relatedness to a particular probe. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, N.Y.; and Ausubel *et al.*, 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Green Publishing Associates and Wiley Interscience, N.Y.

Structural relatedness between two polynucleotide sequences can be expressed as a function of "stringency" of the conditions under which the two sequences will hybridize with one another. As used herein, the term "stringency" refers to the extent that the conditions disfavor hybridization. Stringent conditions strongly disfavor hybridization, and only the most structurally related molecules will hybridize to one another under such conditions. Conversely, non-stringent conditions favor hybridization of molecules displaying a lesser degree of structural relatedness. Hybridization stringency, therefore, directly correlates with the structural relationships of two nucleic acid sequences. The following relationships are useful in correlating hybridization and relatedness (where T_m is the melting temperature of a nucleic acid duplex):

- a. $T_m = 69.3 + 0.41(G+C)\%$
- b. The T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched base pairs.
- c. $(T_m)_{\mu 2} - (T_m)_{\mu 1} = 18.5 \log_{10} \mu 2 / \mu 1$
where $\mu 1$ and $\mu 2$ are the ionic strengths of two solutions.

Hybridization stringency is a function of many factors, including overall DNA concentration, ionic strength, temperature, probe size and the presence of agents which disrupt hydrogen bonding. Factors promoting hybridization include high DNA

concentrations, high ionic strengths, low temperatures, longer probe size and the absence of agents that disrupt hydrogen bonding.

Hybridization usually is done in two stages. First, in the "binding" stage, the probe is bound to the target under conditions favoring hybridization. Stringency is usually controlled at this stage by altering the temperature. For high stringency, the temperature is usually between 65°C and 70°C, unless short (<20 nt) oligonucleotide probes are used. A representative hybridization solution comprises 6X SSC, 0.5% SDS, 5X Denhardt's solution and 100µg of non-specific carrier DNA. See Ausubel *et al.*, *supra*, section 2.9, supplement 27 (1994). Of course many different, yet functionally equivalent, buffer conditions are known. Where the degree of relatedness is lower, a lower temperature may be chosen. Low stringency binding temperatures are between about 25°C and 40°C. Medium stringency is between at least about 40°C to less than about 65°C. High stringency is at least about 65°C.

Second, the excess probe is removed by washing. It is at this stage that more stringent conditions usually are applied. Hence, it is this "washing" stage that is most important in determining relatedness via hybridization. Washing solutions typically contain lower salt concentrations. One exemplary medium stringency solution contains 2X SSC and 0.1% SDS. A high stringency wash solution contains the equivalent (in ionic strength) of less than about 0.2X SSC, with a preferred stringent solution containing about 0.1X SSC. The temperatures associated with various stringencies are the same as discussed above for "binding." The washing solution also typically is replaced a number of times during washing. For example, typical high stringency washing conditions comprise washing twice for 30 minutes at 55° C. and three times for 15 minutes at 60° C.

The present invention includes nucleic acid molecules that hybridize to the inventive molecules under high stringency binding and washing conditions. More preferred molecules (from an mRNA perspective) are those that are at least 50 % of the length of any one of those depicted in below. Particularly preferred molecules are at least 75 % of the length of those molecules.

Substitutions, Insertions, Additions and Deletions

In a general sense, the preferred DNA variants of the invention are those that retain the closest relationship, as described by "sequence identity" to the inventive DNA molecules. According to another aspect of the invention, therefore, substitutions, insertions, additions and deletions of defined properties are contemplated. It will be recognized that sequence

identity between two polynucleotide sequences, as defined herein, generally is determined with reference to the protein coding region of the sequences. Thus, this definition does not at all limit the amount of DNA, such as vector DNA, that may be attached to the molecules described herein. Preferred DNA sequence variants include molecules encoding proteins sharing some or all of any relevant biological activity of the native molecule.

In creating these variants, the skilled worker will be guided by reference to the protein structure. First, insertions and deletions in any recognized functional domain, above, generally should be avoided, except as noted below in the section entitled "Proteins," where this domain is discussed in detail. Alterations in such domains usually will be limited to conservative amino acid substitutions. In addition, where insertions and deletions are desired, this may be accomplished at the N- and/or C-terminus of the protein molecule (or the corresponding coding regions of the DNA). If insertions or deletions are made within the protein, deletions of major structural features usually should be avoided. Thus, a preferred place to make insertion or deletion variants is in non-structural regions, such as linker regions between two alpha helices.

"Substitutions" generally refer to alterations in the DNA sequence which do not change its overall length, but only alter one or more nucleotide positions, substituting one for another in the common sense of the word. One class of preferred substitutions, "degenerate substitutions," are those that do not alter the encoded amino acid sequence. Some substitutions retain 50%, 55%, 60% or 65% identity. Preferred substitutions retain at least about 70% identity, more preferably at least 70% or 75% identity, with the inventive DNAs. Some more preferred molecules have at least about 80% identity, more preferably at least 80% or 85% identity. Particularly preferred DNAs share at least about 90% identity, more preferably at least 90% or 95% identity.

"Insertions," unlike substitutions, alter the overall length of the DNA molecule, and thus sometimes the encoded protein. Insertions add extra nucleotides to the interior (not the 5' or 3' ends) of the subject DNAs. Preferred insertions are made with reference to the protein sequence encoded by the DNA. Thus, it is most preferred to provide an insertion in the DNA at a location that corresponds to an area of the encoded protein which lacks structure. For instance, it typically would not be beneficial, if the preservation of biological activity is desired, to provide an insertion within an alpha-helical region or a beta-pleated sheet. Accordingly, non-structural areas, such as those containing helix-breaking glycines

and proline residues, are most preferred sites of insertion. Other preferred sites of insertion are the splice sites, which are indicated above in the description of the inventive DNA molecules.

While the optimal size of insertions will vary depending upon the site of insertion and its effect on the overall conformation of the encoded protein, some general guides are useful. Generally, the total insertions (irrespective of their number) should not add more than about 30% (or preferably not more than 30%) to the overall size of the encoded protein. More preferably, the insertion adds less than about 10-20% (yet more preferably 10-20%) in size, with less than about 10% being most preferred. The number of insertions is limited only by the number of suitable insertions sites, and secondarily by the foregoing size preferences.

"Additions," like insertions, also add to the overall size of the DNA molecule, and usually the encoded protein. However, instead of being made within the molecule, they are made on the 5' or 3' end, usually corresponding to the N- or C- terminus of the encoded protein. Unlike deletions, additions are not very size-dependent. Indeed, additions may be of virtually any size. Preferred additions, however, do not exceed about 100% of the size of the native molecule. More preferably, they add less than about 60 to 30% to the overall size, with less than about 30% being most preferred.

"Deletions" diminish the overall size of the DNA and, therefore, also reduce the size of the protein encoded by that DNA. Deletions may be made from either end of the molecule or internal to it. Typical preferred deletions remove discrete structural features of the encoded protein. For example, some deletions will comprise the deletion of one or more exons which may define a structural feature. Preferred deletions remove less than about 30% of the size of the subject molecule. More preferred deletions remove less than about 20% and most preferred deletions remove less than about 10%.

Computer-Defined Variants and Definition of "Sequence Identity"

In general, both the DNA and protein molecules of the invention can be defined with reference to "sequence identity." As used herein, "sequence identity" refers to a comparison made between two molecules using, for example, the standard Smith-Waterman algorithm that is well known in the art.

Some molecules have at least about 50%, 55% or 60% identity. Preferred molecules are those having at least about 65% sequence identity, more preferably at least 65% or 70% sequence identity. Other preferred molecules have at least about 80%, more preferably at

least 80% or 85%, sequence identity. Particularly preferred molecules have at least about 90% sequence identity, more preferably at least 90% sequence identity. Most preferred molecules have at least about 95%, more preferably at least 95%, sequence identity. As used herein, two nucleic acid molecules or proteins are said to "share significant sequence identity" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) identity.

"Sequence identity" is defined herein with reference the Blast 2 algorithm, which is available at the NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>), using default parameters. References pertaining to this algorithm include: those found at http://www.ncbi.nlm.nih.gov/BLAST/blast_references.html; Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266-272; Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131-141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402; and Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656.

METHODS OF MAKING VARIANTS

It will be recognized that variants of the inventive molecules can be constructed in several different ways. For example, they may be constructed as completely synthetic DNAs. Methods of efficiently synthesizing oligonucleotides in the range of 20 to about 150 nucleotides are widely available. See Ausubel *et al.*, *supra*, section 2.11, Supplement 21 (1993). Overlapping oligonucleotides may be synthesized and assembled in a fashion first reported by Khorana *et al.*, J. Mol. Biol. 72:209-217 (1971); see also Ausubel *et al.*, Section 8.2. The synthetic DNAs are designed with convenient restriction sites engineered at the 5' and 3' ends of the gene to facilitate cloning into an appropriate vector.

An alternative method of generating variants is to start with one of the inventive DNAs and then to conduct site-directed mutagenesis. See Ausubel *et al.*, *supra*, chapter 8, Supplement 37 (1997). In a typical method, a target DNA is cloned into a single-stranded

DNA bacteriophage vehicle. Single-stranded DNA is isolated and hybridized with a oligonucleotide containing the desired nucleotide alteration(s). The complementary strand is synthesized and the double stranded phage is introduced into a host. Some of the resulting progeny will contain the desired mutant, which can be confirmed using DNA sequencing. In addition, various methods are available that increase the probability that the progeny phage will be the desired mutant. These methods are well known to those in the field and kits are commercially available for generating such mutants.

ISOLATING HOMOLOGS

Methods

By using the sequences disclosed herein as probes or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs. "Homologs" are essentially naturally-occurring variants and include allelic, species-specific and tissue-specific variants.

Region-specific primers or probes derived from the nucleotide sequence(s) provided can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog using known methods (Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)). Such an application is useful in diagnostic methods, as described in more detail below, as well as in preparing full-length DNAs from various sources. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. As a general guide, the formula $3(G+C) + 2(A+T) = ^\circ\text{C}$, is useful.

When using primers derived from the inventive sequences, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C), only sequences with greater than 75% sequence identity to the primer will be amplified. By employing lower stringency conditions (e.g., annealing at 35-37°C), sequences which have greater than 40-50% sequence identity to the primer also will be amplified.

The PCR product may be subcloned and sequenced to confirm that it indeed displays the expected sequence identity. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled

and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see e.g., Sambrook et al., 1989, *supra*.

When using DNA probes derived from the inventive sequences for colony/plaque hybridization, one skilled in the art will recognize that by employing medium to high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPE and 50% formamide, and washing at 50-65°C in 0.5X SSPE), sequences having regions with greater than 90% sequence identity to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPE and 40-45% formamide, and washing at 42°C in SSPE), sequences having regions with greater than 35-45% sequence identity to the probe will be obtained.

Suitably, genomic or cDNA libraries can be constructed and screened in accord with the previous paragraph. The libraries should be derived from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. The clone containing the homolog may then be purified through methods routinely practiced in the art, and subjected to sequence analysis.

Additionally, an expression library can be constructed utilizing DNA isolated from or cDNA synthesized from a tissue or organism that is known to express the gene of interest, or that is suspected of expressing the gene. In this manner, clones may be induced and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal gene product, as described herein. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, *ANTIBODIES: A LABORATORY MANUAL*, Cold Spring Harbor Press, Cold Spring Harbor Press.)

Human Homologs

Any organism or tissue can be used as the source for homologs of the present invention so long as the organism or tissue naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs is human.

PROTEINS OF THE INVENTION

One class of proteins included within the invention is encoded by the inventive DNA molecules presented. Other proteins according to the invention are those encoded by the DNA variants described above. As noted, these variants are designed with the encoded proteins in mind.

A preferred class of protein fragments includes those fragments which retain any biological activity. These molecules share functional features common the family of proteins, although these characteristics may vary in degree.

According to one aspect of the invention fragments of the inventive proteins are contemplated. Some preferred fragments are those which are capable of eliciting an immune response. Generally these "antigenic" fragments will be from about five amino acids in length to about fifty amino acids in length. Some preferred antigenic fragments are from five to about twenty amino acids long. "Antigenic" response may refer to a T cell response, a B cell response or a response by cells of the macrophage/monocyte lineages. In most cases, however, it will refer to the immune response involved in the generation of antibodies. In other words, the relevant immune response is that of helper T cells and/or B cells. These preferred molecules comprise one or more T cell and/or B cell epitopes.

ANTIBODIES OF THE INVENTION

Antibodies raised against the proteins and protein fragments of the invention also are contemplated by the invention. Described below are antibody products and methods for producing antibodies capable of specifically recognizing one or more epitopes of the presently described proteins and their derivatives.

Antibodies include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies including single chain Fv (scFv) fragments, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, epitope-binding fragments, and humanized forms of any of the above.

As known to one in the art, these antibodies may be used, for example, in the detection of a target protein in a biological sample. They also may be utilized as part of treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels or for the presence of abnormal forms of the such proteins.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol. Methods* 35:1-21 (1980); Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983); Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985), pp. 77-96). Antibodies may also be generated by the known techniques of phage display and *in vitro* immunization.

Polyclonal Antibodies

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as an inventive protein or an antigenic derivative thereof.

Polyclonal antiserum, containing antibodies to heterogeneous epitopes of a single protein, can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified, as known in the art, to enhance immunogenicity. Immunization methods include subcutaneous or intraperitoneal injection of the polypeptide.

Effective polyclonal antibody production is affected by many factors related both to the antigen and to the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and/or adjuvant. In addition, host animal response may vary with site of inoculation. Both inadequate or excessive doses of antigen may result in low titer antisera. In general, however, small doses (high ng to low μ g levels) of antigen administered at multiple intradermal sites appears to be most reliable. Host animals may include but are not limited to rabbits, mice, chickens and rats, to name but a few. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al., *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

The protein immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin β -galactosidase) or through the inclusion of an adjuvant during immunization. Adjuvants include Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

Booster injections can be given at regular intervals, with at least one usually being required for optimal antibody production. The antiserum may be harvested when the antibody titer begins to fall. Titer may be determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen. See, for example, Ouchterlony *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μ M). The antiserum may be purified by affinity chromatography using the immobilized immunogen carried on a solid support. Such methods of affinity chromatography are well known in the art.

Affinity of the antisera for the antigen may be determined by preparing competitive binding curves, as described, for example, by Fisher, Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D.C. (1980).

In addition to using protein as the immunogen, DNA molecules may be used directly. In this manner, a DNA encoding the protein immunogen is administered. Boosting and harvesting is done in a manner analogous to that detailed above. Yet another method of producing antibodies entails immunizing chickens and harvesting the antibodies from their eggs.

Monoclonal Antibodies

Monoclonal antibodies (MAbs), are homogeneous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture or *in vivo*. MAbs may be produced

such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvate may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-*p*-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the composition may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for

use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

RECOMBINANT CONSTRUCTS AND EXPRESSION

The present invention further provides recombinant DNA constructs comprising one or more of the nucleotide sequences of the present invention. The recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a DNA or DNA fragment, typically bearing an open reading frame, is inserted, in either orientation.

The gene products encoded by the subject DNAs may be produced by recombinant DNA technology using techniques well known in the art. See, for example, the techniques described in Sambrook et al., 1989, *supra*, and Ausubel et al., 1989, *supra*. Alternatively, the DNA sequences may be chemically synthesized using, for example, synthesizers. See, for

example, the techniques described in OLIGONUCLEOTIDE SYNTHESIS, 1984, Gait, ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety. They may be assembled from fragments and short oligonucleotide linkers, or from a series of oligonucleotides. They are preferably made by RT-PCR methods. The resulting synthetic gene is capable of being expressed in a recombinant vector.

In some cases the recombinant constructs will be expression vectors, which are capable of expressing the RNA and/or protein products of the encoded DNA(s). Thus, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the open reading frame (ORF). The vector may further comprise a selectable marker sequence.

Specific initiation signals may also be required for efficient translation of inserted target gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where a target DNA includes its own initiation codon and adjacent sequences is inserted into the appropriate expression vector, no additional translation control signals may be needed. However, in cases where only a portion of an ORF is used, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire target. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner *et al.*, *Methods in Enzymol.* 153:516-544 (1987)). Some appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U.S. Patent No. 5,082,767.

The present invention further provides host cells containing at least one of the DNAs of the present invention. The host cell can be virtually any cell for which expression vectors are available. It may be, for example, a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic

cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis *et al.*, *Basic Methods in Molecular Biology* (1986)).

A wide variety of expression systems are available, such as: yeast (*e.g.* *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the target DNA; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing the target DNA sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.* Ti plasmid) containing target DNA coding sequences; or mammalian cell systems (*e.g.* COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

Depending on the system chosen, the resulting product may differ. For example, proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

Vectors

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting selection of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequence, and in one aspect of the invention, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal or C-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Bacterial Expression

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, if desirable, to provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may, also be employed as a matter of choice.

Bacterial vectors may be, for example, bacteriophage-, plasmid- or cosmid-based. These vectors can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids typically containing elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, GEM 1 (Promega Biotec, Madison, WI, USA), pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pKK232-8, pDR540, and pRIT5 (Pharmacia).

These "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Bacterial promoters include lac, T3, T7, lambda P_R or P_L, trp, and ara.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is derepressed/induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the coding sequence may be ligated into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye et al. 1985, *Nucleic Acids*

Res. 13:3101-3109; Van Heeke *et al.*, 1989, *J. Biol. Chem.* 264:5503-5509); pET vectors, Studier *et al.*, *Methods in Enzymology* 185: 60-89 (Academic Press 1990); and the like.

Moreover, pGEX vectors may be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and easily can be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a one embodiment, full length cDNA sequences are appended with in-frame *Bam*HI sites at the amino terminus and *Eco*RI sites at the carboxyl terminus using standard PCR methodologies (Innis *et al.*, 1990, *supra*) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.* 1985, *EMBO J.* 4: 1075; Zabeau and Stanley, 1982, *EMBO J.* 1:1217).

Eukaryotic Expression

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Mammalian promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Exemplary mammalian vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). Selectable markers include CAT (chloramphenicol transferase).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the coding sequence of interest

may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a target protein in infected hosts. (*E.g.*, See Logan *et al.*, 1984, *Proc. Natl. Acad. Sci. USA* 81:3655-3659).

In one embodiment, cDNA sequences encoding the full-length open reading frames are ligated into pCMV β replacing the β -galactosidase gene such that cDNA expression is driven by the CMV promoter (Alam, 1990, *Anal. Biochem.* 188: 245-254; MacGregor *et al.*, 1989, *Nucl. Acids Res.* 17: 2365; Norton *et al.* 1985, *Mol. Cell. Biol.* 5: 281).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins.

Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins in eukaryotic cells, stable expression is preferred. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker.

Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the target protein. Such engineered cell lines may be

particularly useful in screening and evaluation of compounds that affect the endogenous activity of the protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, *et al.*, *Cell* 11:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase (Szybalska *et al.*, *Proc. Natl. Acad. Sci. USA* 48:2026 (1962)), and adenine phosphoribosyltransferase (Lowy, *et al.*, *Cell* 22:817 (1980)) genes can be employed in tk⁻, hgp^rt⁻ or apr^t- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, *et al.*, *Proc. Natl. Acad. Sci. USA* 77:3567 (1980)); O'Hare, *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, *J. Mol. Biol.* 150:1); and hydro, which confers resistance to hygromycin (Santerre, *et al.*, 1984, *Gene* 30:147) genes.

An alternative fusion protein system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 8972-8976 (1991)). In this system, the gene of interest is subcloned into a vaccinia-based plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni²⁺ nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The target coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a target gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (E.g., see Smith *et al.*, 1983, *J. Virol.* 46: 584; Smith, U.S. Patent No. 4,215,051).

While the present proteins can be expressed in recombinant systems, as described above, cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Purification of Recombinant Proteins

Recombinant proteins produced may be isolated by host cell lysis. This may be followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, like lysozyme and chelators.

If inclusion bodies are formed in bacterial systems, they may be extracted from cell pellets using, for example, detergents, reducing agents, salts, urea, guanidinium chloride and extremes of pH (*e.g.* <4 or >10). If denaturation occurs, protein refolding steps (*e.g.*, dialysis) can be used, as necessary, in completing configuration of the mature protein. If disulfide bridges are present in the native protein, they may be reoxidized using known methods.

By way of specific non-limiting example, the recombinant bacterial cells, for example *E. coli*, are grown in any of a number of suitable media, for example LB, and the expression of the recombinant protein induced by adding IPTG (*e.g.*, *lac* operator-promoter) to the media or switching incubation to a higher temperature (*e.g.*, λ cl^{857}). After culturing the bacteria for a further period of between 2 and 24 hours, the cells are collected by centrifugation and washed to remove residual media. The bacterial cells are then lysed, for example, by disruption in a cell homogenizer and centrifuged to separate the cell membranes from the soluble cell components. If the protein aggregates into inclusion bodies, this centrifugation can be performed under conditions whereby the dense inclusion bodies are selectively enriched by incorporation of sugars such as sucrose into the buffer and centrifugation at a selective speed. The inclusion bodies can then be washed in any of several solutions to remove some of the contaminating host proteins, then solubilized in solutions containing high concentrations of urea (*e.g.* 8M) or chaotropic agents such as guanidinium hydrochloride in the presence of reducing agents such as β -mercaptoethanol or DTT (dithiothreitol).

At this stage it may be advantageous to incubate the protein for several hours under conditions suitable for the protein to undergo a refolding process into a conformation which

more closely resembles that of the native protein. Such conditions generally include low protein concentrations less than 500 µg/ml), low levels of reducing agent, concentrations of urea less than 2 M and often the presence of reagents such as a mixture of reduced and oxidized glutathione which facilitate the interchange of disulphide bonds within the protein molecule. The refolding process can be monitored, for example, by SDS-PAGE or with antibodies which are specific for the native molecule. Following refolding, the protein can then be purified further and separated from the refolding mixture by chromatography on any of several supports including ion exchange resins, gel permeation resins or on a variety of affinity columns.

Labeling Proteins

When used as a component in assay systems such as those described, below, the target protein may be labeled, either directly or indirectly, to facilitate detection of the present *res*-like molecules either *in vitro* or *in vivo*. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ¹²⁵I; enzyme labeling systems that generate a detectable colorimetric signal or light when exposed to substrate; and fluorescent labels.

Where recombinant DNA technology is used for protein production the, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection. These fusion proteins may, for example, add amino acids which facilitate further chemical modification. They also may add a functional moiety, such as an enzyme, which directly facilitates detection.

TRANSGENIC ANIMALS

The invention further contemplates animal models for studying the function of the present molecules and for overproducing the protein products. The disclosed DNA sequences may be used in conjunction with techniques for producing transgenic animals that are well known to those of skill in the art.

To prepare transgenic animals, target gene sequences may for example be introduced into, and overexpressed in, the genome of the animal of interest, or, if endogenous target gene sequences are present, they may either be overexpressed or, alternatively, be disrupted in order to underexpress or inactivate target gene expression, such as described for the disruption of apoE in mice (Plum *et al.*, *Cell* 71: 343-353 (1992)).

In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the animal and cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation.

For underexpression of an endogenous target gene sequence, such a sequence may be isolated and engineered such that when reintroduced into the genome of the animal of interest, the endogenous target gene alleles will be inactivated. Preferably, the engineered target gene sequence is introduced via gene targeting such that the endogenous target sequence is disrupted upon integration of the engineered target gene sequence into the animal's genome.

Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate cardiovascular disease animal models. Goats, cows and sheep are particularly preferred for producing protein *in vivo*.

Any technique known in the art may be used to introduce a target gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe *et al.*, U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82:6148-6152 (1985)); gene targeting in embryonic stem cells (Thompson *et al.*, *Cell* 56:313-321 (1989)); electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803-1814 (1983)); and sperm-mediated gene transfer (Lavitrano *et al.*, *Cell* 57:717-723 (1989)); *etc.* For a review of such techniques, see Gordon, Transgenic Animals, *Intl. Rev. Cytol.* 115:171-229 (1989).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching

of Lasko et al. (Lasko *et al.*, *Proc. Natl. Acad. Sci. USA* 89:3232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the target gene be integrated into the chromosomal site of the endogenous target gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous target gene of interest are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous target gene.

The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene of interest in only that cell type, by following, for example, the teaching of Gu *et al.* *Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant target gene and protein may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and RT-PCR. Samples of target gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the target gene transgene gene product of interest.

The transgenic animals that express target gene mRNA or target gene transgene peptide (detected immunocytochemically, using antibodies directed against the target gene product's epitopes) at easily detectable levels should then be further evaluated to identify those animals which display characteristic increased susceptibility to carcinogenesis. Additionally, specific cell types within the transgenic animals may be analyzed and assayed *in vitro* for cellular phenotypes characteristic of mutant phenotype.

Once target gene transgenic founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include but are not limited to: outbreeding of founder animals with more

than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound target gene transgenics that express the target gene transgene of interest at higher levels because of the effects of additive expression of each target gene transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order both to augment expression and eliminate the possible need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; breeding animals to different inbred genetic backgrounds so as to examine effects of modifying alleles on expression of the target gene transgene and the possible development of carcinogenesis. One such approach is to cross the target gene transgenic founder animals with a wild type strain to produce an F1 generation that exhibits increased susceptibility to carcinogenesis. The F1 generation may then be inbred in order to develop a homozygous line, if it is found that homozygous target gene transgenic animals are viable.

Methods of generating "knockout" mice using homologous recombination in embryonic stem cells are well known in the art. Suitable methods are described, for example, in Mansour *et al.*, *Nature*, 336:348 (1988); Zijlstra *et al.*, *Nature*, 342:435 (1989) and 344:742 (1990); and Hasty *et al.*, *Nature*, 350:243 (1991). This genomic DNA can be obtained by conventional methods using the cDNA sequence as a probe in a commercially-available genomic DNA library.

Briefly, a genomic fragment is cleaved with a restriction endonuclease and a heterologous cassette containing a neomycin-resistance gene is inserted at the cleavage site. A suitable cassette is the GTI-II *neo* cassette described by Lufkin *et al.*, *Cell* 66:1105 (1991). The modified genomic fragment is cloned into a suitable targeting vector that is introduced into murine embryonic stem cells by electroporation. Cells that have undergone homologous recombination (and hence disruption of the gene) are selected by resistance to G418, and used to generate chimeric mice using well known methods. See Lufkin *et al.*, *supra*. Traditional breeding methods then can be used to generate mice that are homozygous for the disrupted gene.

The phenotype of mice that are homozygous for the mutation then can be studied to provide insights into the role of the protein in, for example, carcinogenesis. These mice also can be used as models for developing new treatments for cancers. If this mutation is lethal in

homozygous mice (for example during embryogenesis) heterozygous mice, which express only half the amount of the protein can also be studied.

GENE THERAPY APPLICATIONS

When mutations in the inventive protein, or in the elements controlling expression of that protein, are found to be associated with a malignant phenotype, control of cellular proliferation can be restored by gene therapy methods. For example, overexpression of the protein can be counteracted by concurrent expression of an antisense molecule that binds to and inhibits expression of the mRNA encoding the protein. Alternatively, overexpression can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. In another embodiment, where expression of a mutated protein induces the malignant phenotype, concomitant expression of the non-mutated molecule via introduction of an exogenous gene may be used. Methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art.

Each of these methods requires a system for introducing a vector into the cells containing the mutated gene. The vector encodes either an antisense or ribozyme transcript of the inventive protein. The construction of a suitable vector can be achieved by any of the methods well-known in the art for the insertion of exogenous DNA into a vector. *See, e.g.,* Sambrook *et al.*, *Molecular Cloning* (Cold Spring Harbor Press 2d ed. 1989), which is incorporated herein by reference. In addition, the prior art teaches various methods of introducing exogenous genes into cells *in vivo*. *See* Rosenberg *et al.*, *Science* 242:1575-1578 (1988) and Wolff *et al.*, *PNAS* 86:9011-9014 (1989), which are incorporated herein by reference. The routes of delivery include systemic administration and administration *in situ*. Well-known techniques include systemic administration with cationic liposomes, and administration *in situ* with viral vectors. Any one of the gene delivery methodologies described in the prior art is suitable for the introduction of a recombinant vector containing an inventive gene according to the invention into a MTX-resistant, transport-deficient cancer cell. A listing of present-day vectors suitable for the purpose of this invention is set forth in Hodgson, *Bio/Technology* 13: 222 (1995), which is incorporated by reference.

For example, liposome-mediated gene transfer is a suitable method for the introduction of a recombinant vector containing an inventive gene according to the invention

into a MTX-resistant, transport-deficient cancer cell. The use of a cationic liposome, such as DC-Chol/DOPE liposome, has been widely documented as an appropriate vehicle to deliver DNA to a wide range of tissues through intravenous injection of DNA/cationic liposome complexes. See Caplen *et al.*, *Nature Med.* 1:39-46 (1995) and Zhu *et al.*, *Science* 261:209-211 (1993), which are herein incorporated by reference. Liposomes transfer genes to the target cells by fusing with the plasma membrane. The entry process is relatively efficient, but once inside the cell, the liposome-DNA complex has no inherent mechanism to deliver the DNA to the nucleus. As such, the most of the lipid and DNA gets shunted to cytoplasmic waste systems and destroyed. The obvious advantage of liposomes as a gene therapy vector is that liposomes contain no proteins, which thus minimizes the potential of host immune responses.

As another example, viral vector-mediated gene transfer is also a suitable method for the introduction of the vector into a target cell. Appropriate viral vectors include adenovirus vectors and adeno-associated virus vectors, retrovirus vectors and herpesvirus vectors.

Adenoviruses are linear, double stranded DNA viruses complexed with core proteins and surrounded by capsid proteins. The common serotypes 2 and 5, which are not associated with any human malignancies, are typically the base vectors. By deleting parts of the virus genome and inserting the desired gene under the control of a constitutive viral promoter, the virus becomes a replication deficient vector capable of transferring the exogenous DNA to differentiated, non-proliferating cells. To enter cells, the adenovirus fibre interacts with specific receptors on the cell surface, and the adenovirus surface proteins interact with the cell surface integrins. The virus penton-cell integrin interaction provides the signal that brings the exogenous gene-containing virus into a cytoplasmic endosome. The adenovirus breaks out of the endosome and moves to the nucleus, the viral capsid falls apart, and the exogenous DNA enters the cell nucleus where it functions, in an epichromosomal fashion, to express the exogenous gene. Detailed discussions of the use of adenoviral vectors for gene therapy can be found in Berkner, *Biotechniques* 6:616-629 (1988) and Trapnell, *Advanced Drug Delivery Rev.* 12:185-199 (1993), which are herein incorporated by reference. Adenovirus-derived vectors, particularly non-replicative adenovirus vectors, are characterized by their ability to accommodate exogenous DNA of 7.5 kB, relative stability, wide host range, low pathogenicity in man, and high titers (10^4 to 10^5 plaque forming units per cell). See Stratford-Perricaudet *et al.*, *PNAS* 89:2581 (1992).

Adeno-associated virus (AAV) vectors also can be used for the present invention. AAV is a linear single-stranded DNA parvovirus that is endogenous to many mammalian species. AAV has a broad host range despite the limitation that AAV is a defective parvovirus which is dependent totally on either adenovirus or herpesvirus for its reproduction *in vivo*. The use of AAV as a vector for the introduction into target cells of exogenous DNA is well-known in the art. *See, e.g., Lebkowski et al., Mole. & Cell. Biol.* 8:3988 (1988), which is incorporated herein by reference. In these vectors, the capsid gene of AAV is replaced by a desired DNA fragment, and transcomplementation of the deleted capsid function is used to create a recombinant virus stock. Upon infection the recombinant virus uncoats in the nucleus and integrates into the host genome.

Another suitable virus-based gene delivery mechanism is retroviral vector-mediated gene transfer. In general, retroviral vectors are well-known in the art. *See Breakfield et al., Mole. Neuro. Biol.* 1:339 (1987) and Shih *et al.*, in *Vaccines* 85: 177 (Cold Spring Harbor Press 1985). A variety of retroviral vectors and retroviral vector-producing cell lines can be used for the present invention. Appropriate retroviral vectors include Moloney Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus. These vectors include replication-competent and replication-defective retroviral vectors. In addition, amphotropic and xenotropic retroviral vectors can be used. In carrying out the invention, retroviral vectors can be introduced to a tumor directly or in the form of free retroviral vector producing-cell lines. Suitable producer cells include fibroblasts, neurons, glial cells, keratinocytes, hepatocytes, connective tissue cells, ependymal cells, chromaffin cells. *See Wolff et al., PNAS* 84:3344 (1989).

Retroviral vectors generally are constructed such that the majority of its structural genes are deleted or replaced by exogenous DNA of interest, and such that the likelihood is reduced that viral proteins will be expressed. *See Bender et al., J. Virol.* 61:1639 (1987) and Armento *et al., J. Virol.* 61:1647 (1987), which are herein incorporated by reference. To facilitate expression of the antisense or ribozyme molecule, of the inventive protein, a retroviral vector employed in the present invention must integrate into the genome of the host cell genome, an event which occurs only in mitotically active cells. The necessity for host cell replication effectively limits retroviral gene expression to tumor cells, which are highly

replicative, and to a few normal tissues. The normal tissue cells theoretically most likely to be transduced by a retroviral vector, therefore, are the endothelial cells that line the blood vessels that supply blood to the tumor. In addition, it is also possible that a retroviral vector would integrate into white blood cells both in the tumor or in the blood circulating through the tumor.

The spread of retroviral vector to normal tissues, however, is limited. The local administration to a tumor of a retroviral vector or retroviral vector producing cells will restrict vector propagation to the local region of the tumor, minimizing transduction, integration, expression and subsequent cytotoxic effect on surrounding cells that are mitotically active.

Both replicatively deficient and replicatively competent retroviral vectors can be used in the invention, subject to their respective advantages and disadvantages. For instance, for tumors that have spread regionally, such as lung cancers, the direct injection of cell lines that produce replication-deficient vectors may not deliver the vector to a large enough area to completely eradicate the tumor, since the vector will be released only from the original producer cells and their progeny, and diffusion is limited. Similar constraints apply to the application of replication deficient vectors to tumors that grow slowly, such as human breast cancers which typically have doubling times of 30 days versus the 24 hours common among human gliomas. The much shortened survival-time of the producer cells, probably no more than 7-14 days in the absence of immunosuppression, limits to only a portion of their replicative cycle the exposure of the tumor cells to the retroviral vector.

The use of replication-defective retroviruses for treating tumors requires producer cells and is limited because each replication-defective retrovirus particle can enter only a single cell and cannot productively infect others thereafter. Because these replication-defective retroviruses cannot spread to other tumor cells, they would be unable to completely penetrate a deep, multilayered tumor *in vivo*. See Markert *et al.*, *Neurosurg.* 77: 590 (1992). The injection of replication-competent retroviral vector particles or a cell line that produces a replication-competent retroviral vector virus may prove to be a more effective therapeutic because a replication competent retroviral vector will establish a productive infection that will transduce cells as long as it persists. Moreover, replicatively competent retroviral vectors may follow the tumor as it metastasizes, carried along and propagated by transduced tumor cells. The risks for complications are greater, with replicatively competent vectors, however.

Such vectors may pose a greater risk than replicatively deficient vectors of transducing normal tissues, for instance. The risks of undesired vector propagation for each type of cancer and affected body area can be weighed against the advantages in the situation of replicatively competent versus replicatively deficient retroviral vector to determine an optimum treatment.

Both amphotropic and xenotropic retroviral vectors may be used in the invention. Amphotropic viruses have a very broad host range that includes most or all mammalian cells, as is well known to the art. Xenotropic viruses can infect all mammalian cell cells except mouse cells. Thus, amphotropic and xenotropic retroviruses from many species, including cows, sheep, pigs, dogs, cats, rats, and mice, *inter alia* can be used to provide retroviral vectors in accordance with the invention, provided the vectors can transfer genes into proliferating human cells *in vivo*.

Clinical trials employing retroviral vector therapy treatment of cancer have been approved in the United States. See Culver, *Clin. Chem.* 40: 510 (1994). Retroviral vector-containing cells have been implanted into brain tumors growing in human patients. See Oldfield *et al.*, *Hum. Gene Ther.* 4: 39 (1993). These retroviral vectors carried the HSV-1 thymidine kinase (HSV-tk) gene into the surrounding brain tumor cells, which conferred sensitivity of the tumor cells to the antiviral drug ganciclovir. Some of the limitations of current retroviral based cancer therapy, as described by Oldfield are: (1) the low titer of virus produced, (2) virus spread is limited to the region surrounding the producer cell implant, (3) possible immune response to the producer cell line, (4) possible insertional mutagenesis and transformation of retroviral infected cells, (5) only a single treatment regimen of pro-drug, ganciclovir, is possible because the "suicide" product kills retrovirally infected cells and producer cells and (6) the bystander effect is limited to cells in direct contact with retrovirally transformed cells. See Bi *et al.*, *Human Gene Therapy* 4: 725 (1993).

Yet another suitable virus-based gene delivery mechanism is herpesvirus vector-mediated gene transfer. While much less is known about the use of herpesvirus vectors, replication-competent HSV-1 viral vectors have been described in the context of antitumor therapy. See Martuza *et al.*, *Science* 252: 854 (1991), which is incorporated herein by reference.

DIAGNOSTIC METHODS

The present invention also contemplates, for certain molecules described below, methods for diagnosis of human disease. In particular, patients can be screened for the occurrence of cancers, or likelihood of occurrence of cancers, associated with mutations in the encoded protein. DNA from tumor tissue obtained from patients suffering from cancer can be isolated and the gene encoding the protein can be sequenced. By examining a number of patients in this manner, mutations in the gene that are associated with a malignant cellular phenotype can be identified. In addition, correlation of the nature of the observed mutations with subsequent observed clinical outcomes allows development of prognostic model for the predicted outcome in a particular patient.

Screening for mutations conveniently can be carried out at the DNA level by use of PCR, although the skilled artisan will be aware that many other well known methods are available for the screening. PCR primers can be selected that flank known mutation sites, and the PCR products can be sequenced to detect the occurrence of the mutation. Alternatively, the 3' residue of one PCR primer can be selected to be a match only for the residue found in the unmutated gene. If the gene is mutated, there will be a mismatch at the 3' end of the primer, and primer extension cannot occur, and no PCR product will be obtained. Alternatively, primer mixtures can be used where the 3' residue of one primer is any nucleotide other than the nonmutated residue. Observation of a PCR product then indicates that a mutation has occurred. Other methods of using, for example, oligonucleotide probes to screen for mutations are described, for example, in U.S. Patent No. 4,871,838, which is herein incorporated by reference in its entirety.

Alternatively, antibodies can be generated that selectively bind either mutated or non-mutated protein. The antibodies then can be used to screen tissue samples for occurrence of mutations in a manner analogous to the DNA-based methods described *supra*.

The diagnostic methods described above can be used not only for diagnosis and for prognosis of existing disease, but may also be used to predict the likelihood of the future occurrence of disease. For example, clinically healthy patients can be screened for mutations in the inventive molecule that correlate with later disease onset. Such mutations may be observed in the heterozygous state in healthy individuals. In such cases a single mutation event can effectively disable proper functioning of the gene and induce a transformed or malignant phenotype. This screening also may be carried out prenatally or neonatally.

DNA molecules according to the invention also are well suited for use in so-called "gene chip" diagnostic applications. Such applications have been developed by, *inter alia*, Synteni and Affymetrix. Briefly, all or part of the DNA molecules of the invention can be used either as a probe to screen a polynucleotide array on a "gene chip," or they may be immobilized on the chip itself and used to identify other polynucleotides via hybridization to the surface of the chip. In this manner, for example, related genes can be identified, or expression patterns of the gene in various tissues can be simultaneously studied. Such gene chips have particular application for diagnosis of disease, or in forensic analysis to detect the presence or absence of an analyte. Suitable chip technology is described for example, in Wodicka *et al.*, *Nature Biotechnology*, 15:1359 (1997) which is hereby incorporated by reference in its entirety, and references cited therein.

PROTEIN-PROTEIN INTERACTIONS

Due to their similarity to certain known proteins, it is anticipated that some of the inventive protein molecules will interact with another class of cellular proteins. This is particularly true of those molecule containing leucine zipper motifs.

Any method suitable for detecting protein-protein interactions can be employed for identifying interacting targets. Among the traditional methods which can be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Utilizing procedures such as these allows for the identification of GAP gene products. Once identified, a GAP protein can be used, in conjunction with standard techniques, to identify its corresponding pathway gene. For example, at least a portion of the amino acid sequence of the pathway gene product can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (see, *e.g.*, Creighton, 1983, *PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES*, W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained can be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for pathway gene sequences. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and for screening are well-known. (See *e.g.*, Ausubel, *supra*, and *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, 1990, Innis *et al.*, eds. Academic Press, Inc., New York).

Additionally, methods can be employed which result in the simultaneous identification of interacting target genes. One method which detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration purposes only and not by way of limitation. One version of this system has been described (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 9578-9582 (1991)) and is commercially available from Clontech (Palo Alto, CA).

Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to a known protein, in this case an inventive protein, and the other contains the activator protein's activation domain fused to an unknown protein (a putative GAP, for instance) that is encoded by a cDNA which has been recombined into this plasmid as part of a cDNA library. The plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, *lacZ*) whose regulatory region contains the transcription activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene, the DNA-binding domain hybrid cannot because it does not provide activation function, and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology can be used to screen activation domain libraries for proteins that interact with a known "bait" gene product. By way of example, and not by way of limitation, gene products known to be involved in TH cell subpopulation-related disorders and/or differentiation, maintenance, and/or effector function of the subpopulations can be used as the bait gene products. Total genomic or cDNA sequences are fused to the DNA encoding on activation domain. This library and a plasmid encoding a hybrid of the bait gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, the bait gene can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

The present invention, thus generally described, will be understood more readily by reference to the following examples, which are provided by way of illustration and are not intended to be limiting of the present invention.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

EXAMPLES

EXAMPLE I: cDNA Library Construction

cDNA library plates and clones originated from five cDNA libraries that were constructed by directional cloning. These are available through the Resource Center (<http://www.rzpd.de>) of the German Genome Project. In particular, the hfbr2 (human fetal brain; RZPD number DKFZp564) and hfkd2 (human fetal kidney; DKFZp566) libraries were generated using the Smart kit (Clontech), except that PCR was carried out with primers that contained uracil residues to permit directional cloning without restriction digestion and ligation, and were complementary with the pAMP1 (LifeTechnologies) cloning sites for directional cloning. The htes3 (human testes; DKFZp434), hute1 (human uterus; DKFZp586) and hmcfl (human mammary carcinoma; DKFZp727) libraries are conventional (Gubler, U., Hoffman, B.J., (1983), A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269), size-selected cDNA libraries. They are cloned into pSPORT1 (LifeTechnologies) via a NotI site which is introduced during reverse transcription downstream of the oligo dT primer and a SalI site that is introduced by the ligation of a adapters. The human mammary carcinoma library was constructed from MCF7 cells.

The cDNA sequences of this application were first identified among the sequences comprising various libraries. Technology has advanced considerably since the first cDNA libraries were made. Many small variations in both chemicals and machinery have been instituted over time, and these have improved both the efficiency and safety of the process. Although the cDNAs could be obtained using an older procedure, the procedure presented in this application is exemplary of one currently being used by persons skilled in the art. For the

purpose of providing an exemplary method, the mRNA isolation and cDNA library construction described here is for the MCF-7 library (DKFZp727) from which the clones named DKFZphmcf1_xxyyxx were obtained.

The human cell line MCF-7 was grown in DMEM supplemented with 10% fetal calf serum until confluency. 3×10^8 cells were harvested with a cell scraper in PBS. Cells were lysed in buffer containing 0.5 % NP-40 to leave the nuclei intact. The debris was pelleted by centrifugation at 15 000 x g for 10 minutes at 4 degrees Celsius. Proteins in the supernatant were degraded in presence of SDS and Proteinase K (30 minutes at 56 degrees Celsius). Precipitation of proteins was done in a Phenol/Chloroform extraction, RNA was precipitated from the aqueous phase with Na-acetate and Ethanol. Polyadenylated messages were isolated using Qiagen Oligotex (QIAGEN, Hilden Germany).

First strand cDNA synthesis was accomplished using an oligo (dT) primer which also contained an NotI restriction site. Second strand synthesis was performed using a combination of DNA polymerase I, *E. coli* ligase and RNase H, followed by the addition of a SalI adaptor to the blunt ended cDNA. The SalI adapted, double-stranded cDNA was then digested with NotI restriction enzyme, and fractionated by size on an agarose gel. DNA of the appropriate size was cut from the gel and cast into a second gel in a 90° angle. After electrophoresis in the second dimension, cDNA of the appropriate size was cut from the gel. The agarose block was broken down with help of gelase. The cDNA was purified with help of two phenol extractions and an ethanol precipitation. The cDNA was ligated into SalI/NotI pre-digested pSport1 vector (LifeTechnologies) and transformed into DH10B bacteria.

The libraries were arrayed into 384-well microtiter plates and spotted on high density nylon membranes for hybridization analysis. Filters and clones are available through the Resource Center. Whole plates were distributed to the sequencing partners of the consortium for systematic sequencing.

EXAMPLE II: Sequencing of cDNA Clones

All clones in the 384-well microtiter plates were sequenced from the 5' end. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on

ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry.

The resulting expressed sequence tag (EST) sequences ("r1 ESTs" = sequenced from 5'-end) were analysed for:

a) the lack of identical matches with known genes.

For this, the EST-sequence was blasted against the cDNA consortiums own database and after that against public databases and (with BLASTn and BLASTx against EMBL/EMBLNEW and assembled ESTs, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings). ESTs which were identical to known genes in more than 100 bp, with less than 2 mismatches, were excluded from further analysis.

b) the presence of an open reading frame

Open reading frames (ORFs) were detected with an tool developed by Munich Information Center for Protein Sequences (MIPS) called ORF-map. ORF-map visualises potential start and stop-codons. If an ORF without a stop codon was detected in a r1-EST, the sequence was processed further.

c) the presence of GC rich sequences

A script developed by MIPS computed the GC-content of the r1-sequence, which should be >40%. Writing similar scripts is within the ordinary skill of one in bioinformatics.

d) the lack of repeat structures

Repeats such as Alu, Line or CA-repeats were detected by blasting (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings) against a repeat-database compiled by MIPS. If a repeat was present within the r1-sequence, the sequence were not processed further.

Novel clones that met all criteria were identified to the sequencers, who then performed 3'-end sequencing of these clones. The resulting 3' ESTs ("s1 ESTs" = sequenced from 3'-end) were checked for

a) the lack of matches with known genes in public databases, and sequences already generated by us.

This was done by blasting against EMBL/EMBLNEW and assembled EST (BLASTn and BLASTx, please refer to EXAMPLE III: Bioinformatics analysis of full length cDNAs, for description and parameter settings).

b) the presence of polyadenylation signals.

Again only clones matching the selection criteria were chosen to be sequenced completely by the sequencers. Clones were selected after the following criteria:

A very good ORF had at least one BLASTx match to other proteins. A "good ORF" should extend to the 3' end and be longer than ~40 codons. If the ORF started in the r1 sequence, in front of the potential start codon, there should not exist too many competing start codons in frame with the ORF start codon and the start should match the Kozak consensus ATG. If the EST sequence was too short to decide according to the potential ORF, and there were only a few or no start codons in the sequence the GC content of the Sequence should be greater than 40%. The r1 sequences needed not contain an polyA-tail at the 3' end. In addition, the results of the blasting against the assembled human ESTs could help in questionable cases to decide whether to stop or to continue. A hit against these ESTs was an indication to go further.

Clones passing the above-described screening were sequenced in full. Sequencing was done preferentially using dye terminator chemistry (ABD or Amersham) on ABI automated DNA sequencers (ABI 377, Applied Biosystems), one partner used EMBL prototype instruments (Arakis) mainly with dye primer chemistry. Primer walking (Strauss et al., 1986, Specific-primer-directed DNA sequencing. *Anal Biochem.* 154, 353-360) was the preferred sequencing strategy because of the lower redundancy possible compared to random shotgun (Messing, J., Crea, R., Seeburg, H.P. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9, 32-39) methods. Walking primers were generally designed using software (e.g. Haas, S., Vingron, M., Poustka, A., Wiemann, S. (1998) Primer design in large-scale sequencing. *Nucleic Acids Res.* 26, 3006-3012, Schwager, C., Wiemann, S., Ansorge, W. (1995) GeneSkipper: integrated software environment for DNA sequence assembly and

alignment. HUGO Genome Digest 2, 8-9) that permitted complete automation of this usually time consuming process and helped in the parallel processing of large numbers of clones.

EXAMPLE III: Bioinformatics analysis of full length cDNAs

Each sequence obtained was compared on nucleotide level in a stepwise manner to sequences in EMBL/EMBLNEW, EMBL-EST, EMBL-STS using the BLASTn algorithm. Basic Local Alignment Search Tool (BLAST, Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S. F. et al (1990) J Mol Biol 215:403-10) is used to search for local sequence alignments. BLAST produces alignments of both nucleotide (BLASTn) and amino acid sequences (BLASTp or BLASTx) to determine sequence similarity. BLAST is especially useful in determining exact matches or in identifying homologs, because of the local nature of the alignments. While it is useful for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment BLAST approach is to look threshold or cut off score set by the user. BLAST looks for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output. Parameter settings for the BLAST-operations (BLASTN 2.0a19MP-WashU) described were: EMBL-EMBLNEW: H=0 V=5 B=5 -filter seg; EMBL-EST: H=0 E=1e-10 B=500 V=500 -filter seg; EMBL-STS: H=0 V=5 B=5.

Search against EMBL/EMBLNEW was done to determine whether the cDNAs are already known, and also to find out whether the cDNAs are encoded by genomic sequences already sequenced and published/submitted to these databases.

Search against EMBL-EST was performed to get a first impression how abundant a particular cDNA would be and to get information on tissue specificity (so-called “electronic Northern-Blot”, e.g. some of the cDNAs derived of the testis library show only hits to ESTs also derived of testis libraries).

The cDNA-sequences were blasted against EMBL-STS to determine STS-sequence-match to the cDNA, thus providing a mapping information to the new cDNA.

The potential protein-sequences were generated automatically by a script searching for the longest open reading frame (ORF) in each of the three forward frames with a minimum length of 90 codons. Next, the automatically generated ORFs were translated into protein sequences. These protein sequences were searched against the non redundant protein data set of PIR/SwissProt/Trembel/Tremblnew (BLASTP 2.0a19MP-WashU, parameter setting: V=7 B=7 H=0 -filter seg). If the script generated more than one ORF, one ORF was chosen manually by the annotater according to the degree of similarity to known proteins, the location of the ORF in the cDNA, the length, the amino acid composition and the content of Prosite-Motifs.

Additionally there was a BLASTx (BLASTX 2.0a19MP-WashU against non redundant protein database comprising PIR/SWISSPROT/TREMBL/TREMBLNEW; parameter-settings were: matrix/home/data/blast/matrix/aa/BLOSUM62 H=0 V=5 B=5 -filter seg) search to find potential frame shift in the complementary cds of the cDNAs and to identify unspliced or partly spliced cDNAs. The protein sequence was then transferred to the PEDANT system, in order to generate additional information on the new proteins. PEDANT (Protein Extraction, Description, and ANalysis Tool, Frishman, D. & Mewes, H.-W. (1997) PEDANTic genome analysis. Trends in Genetics , 13, 415-416) is a platform developed at the Munich Information Center for Protein Sequences (MIPS, Munich, Germany), which incorporates practically all bioinformatics methods important for the functional and structural characterisation of protein sequences. Computational methods used by PEDANT are:

FASTA

Very sensitive protein sequence database searches with estimates of statistical significance. Pearson W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* 183, 63-98.

BLAST2

Very sensitive protein sequence database searches with estimates of statistical significance. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-10.

PREDATOR

High-accuracy secondary structure prediction from single and multiple sequences. Frishman, D. and Argos, P. (1997) 75% accuracy in protein secondary structure prediction. *Proteins*, 27, 329-335. Frishman, D. and Argos, P. (1996) Incorporation of long-distance interactions in a secondary structure prediction algorithm. *Prot. Eng.* 9, 133-142.

STRIDE

Secondary structure assignment from atomic coordinates. Frishman, D. and Argos, P. (1995) Knowledge-based secondary structure assignment. *Proteins* 23, 566-579.

CLUSTALW

Multiple sequence alignment. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.

TMAP

Transmembrane region prediction from multiply aligned sequences. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* 237, 182-192.

ALOM2

Transmembrane region prediction from single sequences. Klein, P., Kanehisa, M., and DeLisi, C. Prediction of protein function from sequence properties: A discriminant analysis of a database. *Biochim. Biophys. Acta* 787, 221-226 (1984). Version 2 by Dr. K. Nakai.

SIGNALP

Signal peptide prediction Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10, 1-6.

SEG

Detection of low complexity regions in protein sequences. Wootton, J.C., Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Computers & Chemistry* 17, 149-163.

COILS

Detection of coiled coils. Lupas, A., M. Van Dyke, and J. Stock, "Predicting Coiled Coils from Protein Sequences." *Science* (1991) 252, 1162-1164.

PROSEARCH

Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M., Smith J.E. (1992) ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. *Biotechniques* 13, 919-921.

BLIMPS

Similarity searches against a database of ungapped blocks. J.C. Wallace and Henikoff S., (1992) PATMAT: a searching and extraction program for sequence, pattern and block queries and databases, *CABIOS* 8, 249-254. Written by Bill Alford.

HMMER

Hidden Markov model software . Sonnhammer E.L.L., Eddy S.R., Durbin R. (1997)
 Pfam: A Comprehensive Database of Protein Families Based on Seed Alignments. *Proteins*
 28, 405-420.

pI

Perl script that returns the amino acid composition, molecular weight, theoretical pI, and expected extinction coefficient of an amino acid sequence. By Fred Lindberg. The parameter-settings were as follows: known3d: score > 100; BLAST: E-value < 10; SCOP: <= 50 Alignments, E-Value < 0.0001; signalp: Y=0.7; untersucht vom N-Terminus her: 50 aa; funcat: E-value < 0.001; BLOCKS: <= 10 hits; BLIMPS: threshold 1100.0; COILS: threshold 0.95; SEG: threshold 20.0; BLAST in report: E-value < 0.001; PIR-KW, superfamilies, EC-Nummern in report: E-value < 0.00001; known3d in report: score > 120

The results of PEDANT analysis, together with the results of the similarity searches, constitute the basis for the structural and functional annotation of the cDNAs and the encoded proteins, as specified below.

EXAMPLE III: CELLULAR LOCALIZATIONS OF GFP-FUSION PROTEINS

Plasmids of cDNA-GFP fusions were transfected into mammalian tissue culture cells and allowed to express the proteins for up to 48 hours. Live cells were imaged at 24 hours and 48 hours after transfection and the localisations recorded. The chart, below, depicts the apparent final cellular localisations of 107 cDNA-GFP fusions.

In order to minimize the possibility of the GFP interfering with protein function and/or localization, two separate populations of cDNAs were generated encoding N-terminal or C-terminal GFP fusions. Clearly this appears to be a crucial strategy, since overall only 56% of the proteins localised to a specific compartment irrespective of the position of the GFP. In the instances where only one fusion localized, the complementary fusion either gave no expression or a nuclear and cytosolic staining - characteristic for GFP alone expression.

Each cDNA in turn was subjected to bioinformatic analysis. Where possible, the potential subcellular localisations of the expressed proteins were determined. This

information was then compared to the actual localisations determined from expression of the GFP-fusion proteins in mammalian cells.

DKFZphfbr2_16c16

group: Cell structure and motility

DKFZphfbr2_16c16.3 encodes a novel 586 amino acid protein with similarity to the human actin binding protein MAYVEN and Drosophila Kelch.

MAVEN is a novel actin binding protein predominantly expressed in brain. Drosophila kelch is involved in the maintenance of ring canal organization during oogenesis. The amino half of the protein including the BTB domain mediates dimerization, while the amino half might allow cross-linking of ring canal actin filaments, thus organising the inner rim cytoskeleton. The kelch repeat domain is necessary for ring canal localisation and believed to mediate an additional interaction, possibly with actin. The new protein shares the features of both proteins and therefore should be involved in the organisation of cyto skeleton binding to membrane proteins.

The new protein can find application in modulating/blocking of cyto skeleton-membrane protein interaction.

similarity to Drosophila kelch

complete cDNA, complete cds, EST hits
on genomic level partly encoded by AC005082 and AC006039

Sequenced by Qiagen

Locus: unknown

Insert length: 3028 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2984

```

1  GGGGGCCCGG  GGACGCAGCC  CAGTTGGTAG  CGTCGCTCCC  TGAGCGTTTC
51  TAAGGGGGCC  GCCCGGCCCT  GTCITTCGGC  AGTGGCCGAG  CCACGCCGCG
101 CTGCCGCGCG  TTCCAGAGCT  GGGCGCTGCA  GCTGCACTGC  CGATCGCCGT
151 GTTTGGTTCGA  TAGAATCCCC  AGTGTGCCCA  GAGAGTGCGA  CCCCTCGCCC
201 GGCCCGGCGA  GCCCGGGGCG  TGAACCGAGC  TGAGGGAGGA  TGGCAGCCCT
251 TGGGGTGGAG  AAGAGCAGCA  AGAAGAAGAC  CGAGAAGAAA  CTGCTGCTC
301 GGAAGAAGC  TAAATTGTTG  GCGGGTTTCA  TGGGCGTCAT  GAATACATG
351 CGGAAACAGA  AAACGTTGTG  TGACGTGATC  CTCATGGTCC  AGGAAAGAAA
401 GATACCTGCT  CATCGTGTG  TTCTTGCTGC  AGCCAGTCAT  TTTTAACT
451 TAATGTTTAC  AACTAACATG  CTTGAATCAA  AGTCCTTTGA  AGTAGAATC
501 AAAGATGCTG  AACCTGATAT  TATTGAACAA  CTGGTGGAAT  TTGCTTATAC
551 TGTCTAGAAT  TCCGTGAATA  GCAACAATGT  TCAGTCTTTG  TTGGATGCAG
601 CAAACCAATA  TCAGATTGAA  CCTGTGAAGA  AAATGTGTGT  TGATTTTTTG
651 AAAGAACAAG  TTGATGCTTC  AAATGTCTT  GGTATAAGTG  TGCTAGCGGA
701 GTGCTAGAT  TGTCTGAAT  TGAAGCAAC  TGCAGATGAC  TTTATTCATC
751 AGCACTTTAC  TGAAGTTTAC  AAACTGATG  AATTTCTTCA  ACTTGATGTC
801 AAGCGAGTAA  CACATCTTCT  CAACCAGGAC  ACTCTGACTG  TGAGAGCAGA
851 GGATCAGGTT  TATGATGCTG  CAGTCAGGTG  GTTGAATAAC  GATGAGCCTA
901 ATCGCCAGCC  ATTTATGGTT  GATATCCTTG  CTAAAGTCAG  GTTTCCTCTT
951 ATATCAAGA  ATTTCTTAAG  TAAACCGTA  CAAGCTGAAC  CACTTATTCA
1001 AGACAATCCT  GAATGCCTTA  AGATGGTGAT  AAGTGGAATG  AGGTACCATC
1051 TACTGTCTCC  AGAGGACCGA  GAAGAACTTG  TAGATGGCAC  AAGACCTAGA
1101 AGAAGAAAC  ATGACTACCG  CATAGCCCTA  TTTGGAGGCT  CTCAACCACA
1151 GTCTTGTA  TATTTTAACC  CAAAGGATTA  TAGCTGGACA  GACATCCGCT
1201 GCCCCTTGA  AAAACGAAGA  GATGCAGCAT  GCGTGTTTG  GGACAATGTA
1251 GTATACATT  TGGGAGGCTC  TCAGCTTTTC  CCAATAAAGC  GAATGGACTG
1301 CTATAATGTA  GTGAAGGATA  GCTGGTATTC  GAAACTGGGT  CCTCCGACAC
1351 CTCGAGACAG  CCTTGCTGCA  TGTGCTGCAG  AAGGCAAAAT  TTATACATCT
1401 GGAGGTTTCA  AAGTAGGAAA  CTCAGCTCTG  TATTTATTG  AGTGCTATGA
1451 TACGAGAACT  GAAAGCTGGC  ACACAAAGCC  CAGCATGCTG  ACCCAGCGCT
1501 GCAGCCATGG  GATGGTGGAA  GCCAATGGCC  TAATCTATGT  TTGTGGTGGA
1551 AGTTTAGGAA  ACAATGTTT  AGGGAGAGTG  CTTAATTCCT  GTGAAGTTTA
1601 TGATCCTGCC  ACAGAAACAT  GGACTGAGCT  GTGTCCAATG  ATTGAAGCCA
1651 GGAAGAATCA  TGGGCTGGTA  TTTGTAAAG  ACAAGATATT  TGCTGTGGGT
1701 GGTCAGAATG  GTTTAGGTGG  TCTGGACAAT  GTGGAATATT  ACGATATTAA
1751 GTTGAACGAA  TGAAGATGG  TCTACCAAT  GCCATGGAAG  GGTGTAACAG
1801 TGAATGTGC  AGCAGTTGGC  TCTATAGTTT  ATGCTTGGC  TGGTTTTCAG

```



```

2301 AGAAGATTGG CTCATCAGTG AAGCGCAGTA TCTTAGCTCT AGATTCTATT
2351 TTCATGCATC ACAGAAGTGC TATACGGTTA GGTCTGTTTG TGCTCAGTCA
2401 AGAACTAAGA AATAGTATGA ATTGTAAGTC AAGATGGGCA ACTCAGATGG
2451 AGCAGCTTAG TCTCACAGTT TGCTTGCTCTA TTTATTTTAT TTAGTGCCAA
2501 ATGTATTCCA TTTTAAAAGT AAGCCAGAGT GAGTCAAGGC ATATACACAC
2551 TTTCTCACAA AACTTCCTAA ACAGATTGCG GGGTTTAATA TGTCACATC
2601 CTCATGAAAT ATATTCAATC CACTTAAATA TATTCCATCT TTTTAAACATA
2651 AAATGTAAAG CTTAGCACCC ATCATTAAAT TATGTCTCTG TTTTATCCAG
2701 TGGTAAAAAA AGGATTCTGC CTCTTTAGTC CTCACTGTTA AATAAAACCC
2751 AATCATAGTA AGTGATTAAAC TAGCAAAAAG TAAAGCTATT TATAGCAAAAT
2801 TTCTAGATCA TTAGAAAAGC ACTGGTAGTT GTACAATATC AGTGTGACT
2851 TTGAACCTCT TTAACGAGAT CATGAATTCT TTCCCTTAG CAAAACATG
2901 AAATATTTAA CCTAGTTGTC TCTAAAAGTT TTGTAATCAT GAGTTAGATA
2951 TATGTCATCT CCTATTCAAT GCTTTTATGT GATCAATAAA TCTTTACAA
3001 ACCCAAAAGA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry AC005082 from database EMBL:
Homo sapiens clone RG271G13; HTGS phase 1, 7 unordered pieces.
Score = 6460, P = 0.0e+00, identities = 1292/1292
4 exons matching Bp 1180-3007

Entry AC006039 from database EMBL:
*** SEQUENCING IN PROGRESS *** Homo sapiens clone NH0319F03; HTGS phase
1, 3 unordered pieces.
Score = 1780, P = 2.0e-117, identities = 368/377
5 exons matching Bp 6-860

Entry HSG20603 from database EMBL:
human STS A005Y34.
Score = 670, P = 1.0e-23, identities = 134/134

Medline entries

93201592:
kelch encodes a component of intercellular bridges in
Drosophila egg chambers.

97412177:
Drosophila kelch is an oligomeric ring canal actin organizer.

Peptide information for frame 3

ORF from 240 bp to 1997 bp; peptide length: 586
Category: strong similarity to known protein

```

1 MAASGVEKSS KKKTEKKLAA REEAKLLAGF MGVMMNMRKQ KTLCDVILMV
51 QERKIPAHRV VLAAASHFFN LMFTTNMLES KSFEVELKDA EPDIEQLVE
101 FAYTARISVN SNNVQSLLDA ANQYQIEPVK KMCVDFLKEQ VDASNCLGIS
151 VLAECCLDCPE LKATADDFIH QHFTEVYKTD EFLQLDVKRV THLLNQDTLT
201 VRAEDQVYDA AVRWLKYDEP NRQPFMVIL AKVRFPLISK NFLSKTVQAE
251 PLIQDNPECL KMVISGMRYH LLSPEDEREL VDGTRPRRKK HDYRIALFGG
301 SQPQSCRYFN PKDYSWTDIR CPFKEKRRDA CVFWDNVVYI LGGSQLEFIK
351 RMDCYNVVKD SWYSKLGPPPT PRDLSAACAA EGKIYTSGLS EVGNSALYLF
401 ECYDTRTESW HTKPSMLTQR CSHGMVEANG LIYVCGGSLG NNVSGRVLNS
451 CEVYDPATET WTELCPMIEA RKNHGLVFVK DKIFAVGGQN GLGGLDNVEY
501 YDIKLNEWKM VSPMPWKGVV VKCAAVGSIV YVLAFQGVG RLGHILEYNT
551 ETDKWVANSK VRAFPVTSCL ICVVDTCGAN EETLET

```

BLASTP hits

Entry KELC_DROME from database SWISSPROT:
RING CANAL PROTEIN (KELCH PROTEIN).
Length = 689
Score = 816 (287.2 bits), Expect = 1.9e-81, P = 1.9e-81
Identities = 187/542 (34%), Positives = 290/542 (53%)

Entry AC004021_1 from database TREMBL:
WUGSC:H DJ0186K10.1"; Human PAC clone DJ0186K10 from 5q31,
complete sequence. Homo sapiens (human)
Length = 497

Entry A45773 from database PIR:
kelch protein, long form - fruit fly (*Drosophila melanogaster*)
Length = 1476
Score = 817 (287.6 bits), Expect = 1.7e-80, P = 1.7e-80
Identities = 189/549 (34%), Positives = 292/549 (53%)

```

SEQ  CSHGMVEANGLIYVCGGSLGNNVSGRVLNSCEVYDPATETWTELCPMIEARKNHGLVFK
SEG  .....
PRD  cccceeeccceeecccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  DKIFAVGGQNGGLDNVEYYDIKLNEWKMVSFMPWKGVTVKCAAVGSIVYVLAFQGVG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  RLGHILEYNTETDKWVANSKVRAFPVTSCLICVVDTCGANEETLET
SEG  .....
PRD  cccceeecccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_16c16.3

PS00001	442->446	ASN_GLYCOSYLATION	PDOC00001
PS00004	11->15	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	188->192	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	552->555	PKC_PHOSPHO_SITE	PDOC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDOC00006
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	315->319	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	405->409	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	550->554	CK2_PHOSPHO_SITE	PDOC00006
PS00007	202->209	TYR_PHOSPHO_SITE	PDOC00007
PS00008	5->11	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	389->395	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	436->442	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00008	493->499	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16c16.3)

DKFZphfbr2_16f21

group: brain derived

DKFZphfbr2_16f21 encodes a novel 208 amino acid protein with strong similarity to human zinc finger protein 216.

The novel protein shows strong similarity to the human zinc finger protein 216, but has no Zn finger.

PROSITE: Contains no Zinc finger; No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to zinc finger protein 216

complete cDNA, complete cds, EST hits
start matches Kozak consensus ANNatgG,

Sequenced by Qiagen

Locus: unknown

Insert length: 1512 bp

Poly A stretch at pos. 1490, polyadenylation signal at pos. 1474

```
1 GGGAGCAAGC AGGGGTTCCG CGGCATTACC TGTACCCATT CACCGGCGGC
51 TACCGGCGGC GCGCGGTAGC GTGTCAGGCG GAGAGACCCG CCGCCAGGTG
101 TGCAACTGAG GAACATGGCT CAAGAACTA ATCAGACCCA AGTGCCTATG
151 CTTTGTTCCT CTGGCTGTGG ATTTTATGGA AACCCCTCGTA CAAATGGCAT
201 GTGTTCACTA TGCTATAAAG AACATCTTCA AAGACAGAAT AGTAGTAATG
251 GTAGAATAAG CCCACCTGCA ACCTCTGTCA GTAGTCTGTC TGAATCTTTA
301 CCAGTTCAAT GCACAGATGG CAGTGTGCCA GAAGCCAGT CAGCATTAGA
351 CTCTACATCT TCATCTATGC AGCCAGCCCC TGTATCAAAAT CAGTCACTTT
401 TATCAGAATC TGTAGCATCT TCTCAATTGG ACAGTACATC TGTGGACAAA
451 GCAGTACCTG AAACAGAAGA TGTGCAGGCT TCAGTATCAG ACACAGCACA
501 GCAGCCATCT GAAGAGCAAA GCAAGCCTCT TGA AAAACCG AAACAAAAAA
551 AGAATCGCTG TTTCATGTGC AGGAAGAAAG TGGGACTTAC TGGGTTTGAA
601 TGCCGGTGTG GAAATGTTTA CTGTGGTGTA CACCGTTACT CAGATGTACT
651 CAATTGCTCT TACAATTACA AAGCCGATGC TGCTGAGAAA ATCAGAAAAG
701 AAAATCCAGT AGTTGTTGGT GAAAAGATCC AAAAGATTG AACTCCTGCT
751 GGAATACAAA ATTCTTGAGC ATCTGCAAAAC TAAAAATTGA CTTGAGGTTT
801 TTTTTTCTCT AGTCATTGGG AATGTAGAGC AGTGTATCTT GCATGTCATC
851 GGAAGAATAG ATTTTGTGTT TGGTTTGTGTT TTGAAAATGA CTCTGAACAT
901 TTATTTCCAT TGCAATTCTT GTGGCTGAGG AGACTTAAAC TTACAAGTA
951 TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT
1001 AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTTCCTA GTATTAAACA
1051 TGCATGCATT AATCTTGAGC TTTATTTTCT CATTATGTAT GTATATATCG
1101 CTTTTCTCTG CAGCACGATT TCTCTTTTGA TAATGCCCTT TAGGGCACAA
1151 CTAGTTATCA GTAACATGAT GTATCTTAAT CATTATGGCT GCTTCTGTTT
1201 TTTTCATTAA AAAGGTTATT CATATGTTAG CATATAGTTT CTTTGACCCC
1251 ACTATTTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT
1301 TGTAAAGTTG TGTGTTTAAA CTTTTTTTGG AGCGAGGGAA GAAAAAGCTG
1351 TATGCATTTT ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT
1401 TTGTGTGTAT ACAATATGAA GAATGATCTG AAGTAATTGT GCTGTATTTA
1451 TGTATTATCA CAGTCTTTG ATTAATAAAA AAGGAAAACC AGAAAAAAA
1501 AAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 115 bp to 738 bp; peptide length: 208
 Category: strong similarity to known protein

```

1 MAQETNHSQV PMLCSTGCGF YGNPRTNGMC SVCYKEHLQR QNSSNGRISP
51 PATSVSSLSE SLPVQCTDGS VPEAQSALDS TSSSMQPSPV SNQSLLESV
101 ASSQLDSTSV DKAVPETEDV QASVSDTAQQ PSEEQSKPLE KPKQKKNRCF
151 MCRKKVGLTG FECRCGNVYC GVHRYSDVLN CSYNYKADAA EKIRKENPVV
201 VGEKIQKI

```

BLASTP hits

Entry ATF7H19_1 from database TREMBLNEW:
 gene: "F7H19.10"; product: "putative protein"; Arabidopsis thaliana DNA
 chromosome 4, BAC clone F7H19 (ESSAII project) >TREMBL:ATT12H17_21
 gene: "T12H17.210"; product: "predicted protein"; Arabidopsis thaliana
 DNA chromosome 4, BAC clone T12H17 (ESSAII project)
 Score = 206, P = 2.1e-24, identities = 51/146, positives = 77/146

Entry PVPVPR3A_1 from database TREMBL:
 gene: "PVPVPR3"; P.vulgaris PVPVPR3 protein mRNA, complete cds.
 Score = 237, P = 4.9e-20, identities = 50/136, positives = 73/136

Entry AF062072_1 from database TREMBL:
 gene: "ZNF216"; product: "zinc finger protein 216"; Homo sapiens zinc
 finger protein 216 (ZNF216) gene, complete cds.
 Score = 591, P = 1.6e-57, identities = 124/215, positives = 147/215

Alert BLASTP hits for DKFZphfbr2_16f21, frame 1

TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus
 zinc finger protein ZNF216 mRNA, complete cds., N = 1, Score = 590, P =
 2.1e-57

TREMBLNEW:AB001773_1 gene: "pem-6"; product: "PEM-6"; Ciona savignyi
 pem-6 (posterior end mark 6) mRNA, complete cds., N = 1, Score = 421, P
 = 1.7e-39

>TREMBL:AF062071_1 product: "zinc finger protein ZNF216"; Mus musculus zinc
 finger protein ZNF216 mRNA, complete cds.
 Length = 213

HSPs:

Score = 590 (88.5 bits), Expect = 2.1e-57, P = 2.1e-57
 Identities = 123/213 (57%), Positives = 146/213 (68%)

```

Query:      1 MAQETNHSQV PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPAT---SVSS 57
            MAQETN + PMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQ +S GR+SP T S S
Sbjct:      1 MAQETNQT PGMPLCSTGCGFYGNPRTNGMCSVCYKEHLQRQNS-GRMSPMGTAAGSNSP 59

Query:      58 LSES L P V Q C T D G S V P E A Q S A L D S T S S S M Q P S P V S N Q S L L S E --S V A S S Q L D S T S V D K A V P 115
            +S S V Q D + + A S T S + P V + + + ++ S + D + K
Sbjct:      60 T S D S A S V Q R A D A G L N N C E G A A G S T S E K S R N V P V A A L P V T Q Q M T E M S I S R E D K I T T P K T - E 118

Query:      116 E T E D V Q A S V S D T A Q Q P S E E Q S --K P L E K P K Q K K N R C F M C R K K V G L T G F E C R C G N V Y C G V H 173
            +E V S + Q P S Q S K E P K K N R C F M C R K K V G L T G F + C R C G N + + C G + H
Sbjct:      119 V S E P V V T Q P S P S V S Q P S S S Q S E E K A P E L P K P K K N R C F M C R K K V G L T G F D C R C G N L F C G L H 178

Query:      174 R Y S D V L N C S Y N Y K A D A A E K I R K E N P V V V G E K I Q K I 208
            R Y S D N C Y + Y K A + A A K I R K E N P V V V E K I Q + I
Sbjct:      179 R Y S D K H N C P Y D Y K A E A A A K I R K E N P V V V A E K I Q R I 213

```

Pedant information for DKFZphfbr2_16f21, frame 1

Report for DKFZphfbr2_16f21.1

```

[LENGTH]      208
[MW]           22541.23
[pI]           6.80
[HOMOL]        TREMBL:AF062072_1 gene: "ZNF216"; product: "zinc finger protein 216"; Homo
sapiens zinc finger protein 216 (ZNF216) gene, complete cds. 9e-57
[PIRKW]        zinc 8e-13
[PIRKW]        zinc finger 8e-13

```

```

[PIRKW]      fusion protein 8e-13
[SUPFAM]     unassigned ubiquitin-related proteins 8e-13
[SUPFAM]     ubiquitin homology 8e-13
[PROSITE]    MYRISTYL      2
[PROSITE]    CK2_PHOSPHO_SITE      7
[PROSITE]    ASN_GLYCOSYLATION      4
[KW]         Irregular
[KW]         LOW_COMPLEXITY      7.21 %

SEQ  MAQETNHSQVPMCLSTGCGFYGNPRTNGMCSVCYKEHLQRQNSSNGRISPPATSVSSLSE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccchhhhhhhhhhhccccccccccccccccccccc

SEQ  SLPVQCTDGSVPEAQSALDSTSSSMQSPVSNQSLLESVASSQLDSTSVDKAVPETEDV
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QASVSDTAQQPSEEQSKPLEKPKQKKNRCFCMRKKVGLTGFECRCGNVYCGVHRYSDVLN
SEG  .....
PRD  cccccccccccccccccccccccccccccccccceccccccccceccccccccccccccccccc

SEQ  CSYNYKADAAEKIRKENPVVVGEEKIQKI
SEG  .....
PRD  ccchhhhhhhhhhhhhcccccccccccccc

```

Prosites for DKFZphfbr2_16f21.1

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	42->46	ASN_GLYCOSYLATION	PDOC00001
PS00001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS00001	180->184	ASN_GLYCOSYLATION	PDOC00001
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	123->127	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	166->172	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16f21.1)

DKF2phfbr2_16g18

group: cell cycle

DKF2phfbr2_16g18.3 encodes a novel 984 amino acid protein with similarity to centromeric proteins of yeasts.

The novel protein shows similarity to *S. pombe* SPAC17A5.07c and the *S. cerevisiae* Smt4p suppressor of MIF2 gene. MIF2 encodes a centromeric protein with homology to the mammalian centromeric protein CENP-C. Mutations in MIF2 stabilise dicentric minichromosomes and confer high instability to chromosomes that bear a cis-acting mutation in element I of the yeast centromeric DNA (CDEI). Therefore the new protein should be involved in centromer organisation, too.

The new protein can find application in modulating/blocking the cell cycle and influencing the behavior of chromosomes, both natural and artificial in eukaryotic cells.

similarity to KIAA0797 and yeast Smt4p

complete cDNA, complete cds, EST hits
the yeast Smt4 protein seems to be involved in centromer function
and microtubule organisation

Sequenced by Qiagen

Locus: unknown

Insert length: 4826 bp

Poly A stretch at pos. 4756, polyadenylation signal at pos. 4736

```
1 GGGTCGAGGT CGACGGTATC GATAAGTTTT TTTTTTTTTT TTTTTTTTTT
51 TTTTCTTTTC CCCTCCCCTT CCCTCTCCAA GCCGGAGGGG TCCTGAGGTG
101 ACAGCCGCTG CAATGAAAT TTCAGCAGCG GGAGAAGATG GACACAGAGAA
151 AGCTCGGGCG ACGGCCATCT TCATCCGAAA TCATCACAGA AGGAAAAAGG
201 AAAAAAGTCAT CTTCTGATTT ATCGGAGATA AGAAGATGT TAAATGCAAA
251 ACCAGAGGAT GTCCATGTTT AATCACCCT GTCCAAATTC AGAAGCTCAG
301 AACGCTGGAC TCTCCCTTTG CAGTGGGAAA GAAGCCTAAG GAATAAAGTC
351 ATCTCTCTAG ACCATAAAAA TAAAAACAT ATCCGAGGGT GTCCGTGTAC
401 TTCCAGGTCA TCACCAGAAA GGATACCCAG AGTTATATTG ACGAATGTCC
451 TGGGAACGGA GTTAGGAAGA AAATACATAA GGACCCACCC TGTAAGTCTG
501 GGAAGTTTGA GTGATACAGA CAACCTGCAA TCAGAGCAAC TTTCTTCATC
551 ATCTGATGGC AGCCTAGAAT CTTATCAAAA TCTAAACCCCT CACAAGAGCT
601 GTTATTATTG TGAAAGGGGC TCACAACGAA GTAAAGCAGT AGATGACAAT
651 TCTGCAAAAGC AGACTGCGCA CAATAAAGAA AAACGAAGAA AGGATGATGG
701 CATTTCTCTT TTAATATCTG ATACTCAGCC TGAAGACCTT AACAGTGGAA
751 GTAGAGGTTG TGATCATCTC GAACAGGAAA GCAGAAACAA GGATGTTAAA
801 TATTCTGATT CAAAAGTGA ACTCACTCTG ATTTCCAGGA AGACAAGAG
851 AAGGCTTAGA AATAATTAC CTGATTCTCA ATATTGTACT TCTTTGGATA
901 AGTCAACAGA ACAGACAAAA AAACAAGAAG ATGACTCAAC AATATCCACT
951 GAGTTTGAAA GGCCAAGTGA AAATATCAT CAGGATCCAA AACTGCCTGA
1001 AGAAATTACA ACTAAACCTA CAAAAGTGA TTTTACTAAG CTATCCTCAC
1051 TTAACAGTCA GGAGTTGACT TTGAGTAATG CCACCAAAGG TGCCTCTGCC
1101 GGTTCACCCA CTGAAACCGT TGAGTACTCT AATTCCATTG ATATTGTGGG
1151 GATTCTCTCC CTGGTTGAGA AGGATGAGAA TGAGTTGAAT ACCATAGAAA
1201 AGCCTATTCT AAGAGGACAT AATGAAGGGA ACCAATCACT GATCTCAGCT
1251 GAACCAATPG TTGTTCCAG TGATGAAGAA GGACCTGTGT AACATAAAG
1301 TTCAGAAATT CTTAAGTTAC AATCTAAGCA AGACCGTGAG ACAACTAATG
1351 AAAATGAGAG TACTTCTGAA TCAGCATTGT TAGAACTACC ATTGATTACA
1401 TGTGAATCTG TACAGATGTC ATCTGAATTA TGCCCATATA ATCCTGTCAT
1451 GGAGAACATT TCCAGTATTA TGCCTAGTAA TGAGATGGAT CTACAAGTGG
1501 ATTTTATATT TACTTCTGTT TATATTGGTA AAATAAAGG AGCTTCTAAA
1551 GGTGTGTGTT CAATCACAAA AAAATATATT AAGATCCCAT TTCAAGTGTC
1601 CCTGAATGAG ATTTCAATGC TAGTGGATAC CACACATTTA AAGCGGTTTG
1651 GGTATGGAAG AAGTAAGGAT GATAATCACA GTAAAAGGAG TCATGCTATT
1701 CTTTCTCTCT GGGTCTCTTC AGATTATCTT CAAGAGATTC AGACCCAATT
1751 AGAACACTCT GTATTAGCC AGCAATCAAA ATCTAGTGAA TTCATTTTCC
1801 TTGAACATA CAATCCTGTT TCACAGAGAG AAGAAATTGAA GCTGAAAGAT
1851 ATTATGACGG AAATAAGTAT AATCAGTGGA GAATTAGAGC TTTCTTACCC
1901 GTTGCTTGGG GTTCAGGCAT TTCCTTTGTT TCAGAACCTC TCTTCAAAAG
1951 AAAGTTCTTT TATTCAATTAT TACTGTGTTT CAACTTGTTC TTTCCCTGCT
2001 GGTGTGTGCT TTGCTGAAGA AATGAAGCTG AAATCAGTAT CTCAGCCCTC
2051 AAACACAGAT GCGGCCAAGC CTACTTACAC CTTCTGTCAG AAGCAAAGTA
2101 GCGGTTGCTA CTCCTTTCTT ATTACATCTA ATCCAGATGA AGAATGGCGG
2151 GAAGTCAGGC AACTGGGACT TGTTCAGAGG TTGATTGTAT ATCCTCCACC
2201 ACCCTACTAG GGGGATTTGG GAGTAACATA TGAAGATCTG GAGTGTTTAG
2251 AAGAAGGAGA GTTCTTAAAT GATGTAATCA TTGATTTTTA CCTTAAGTAT
2301 CTTATATTGG AGAAGGCATC AGATGAACCT GTTGAACGAA GTCACATTTT
```

2351 TAGTAGCTTT TTCTATAAAT GCTTGACAAG AAAGGAAAAT AATTTAACAG
2401 AAGATAATCC AAATCTTTCA ATGGCACAGA GAAGACATAA AAGAGTAAGA
2451 ACATGGACTC GTCACATAAA CATTTTTAAT AAAGATTACA TCTTTGTACC
2501 TGTAATAGAG TCGTCTCACT GGTATCTCGC AGTCATTGTG TTCCATGGT
2551 TAGAAGAAGC TGTGTATGAA GATTTTCCAC AAACGTGTATC CCAGCAGTCC
2601 CAGGCTCAGC AGTCCCAAAG TGACAACAAA ACAATAGATA ATGATCTACG
2651 TACTACTTCG ACACGTCTTT TGAGTGCAGA GGATTCCTAA AGTACCGAGT
2701 CGAATATGTC AGTACCAAAG AAAATGTGTA AAAGGCCATG TATTCTTATA
2751 CTAGACTCCT TGAAGCTGCG TTCTGTACGA AACACAGTTC AGAATTTACG
2801 AGAGTATTTA GAGGTAGAGT GGGAAAGTTAA ACTAAAACT CATCGTCAAT
2851 TCAGCAAAC AAACATGGTG GATCTATGCC CTAAGTTC TAAACAGGAC
2901 AATAGCAGTG ATTGTGGAGT ATATTTATTG CAGTATGTGG AAAGCTTCTT
2951 CAAGGATCCT ATTGTAACT TTGAACCTCC AATTCATTG GAGAAGTGGT
3001 TTCCCTCGTCA TGTAATAAAG ACCAAACGGG AAGATATTCG AGAGCTCATC
3051 TTGAACCTTC ATTTACAGCA ACAGAAGGGC AGCAGTAGCT AGTTAATCTG
3101 TACAAACATG ACACAGATGT TCTCTAAGAT TACTGGAAAG CCCCTTACCA
3151 GCATTTGTGT TAGCCAGCTC ACAGAGAAGA AAATAACTTG CAGTAGTTTT
3201 ATAATAAGTC ATTGGAACAT TATTTAAAT ATGTAGGACA CATTATTAGA
3251 ATTGTTGGGA TCTCATAGAT GGAATGGGAA TGGGGGTGAT ATAGATAAAC
3301 TTACTAGATA TAAATTAAAA TTTTATAAAT ATTTTCATATT TTTCTGAGTA
3351 AATATGATTG GATTATGCAA CAGCATATGT AATATGGGAA TGTTTTGTAG
3401 ATAATAAAAC TTACATGATC TGTACTTCCA CGTGACTGGG TGCTGAGGGG
3451 AGTTAAAGCC TCCCTGGTGC CAGCCCCAGT GCTTGTCAA TTGTGTGACA
3501 GGTCACATCA TATTGTAATT CTATTCCTTG CAGCTCAAGC ATGCAGTATG
3551 AATACTGTGT ATTTTTTAAA AAAATAATTT AGTATCAAGG CTTCAGAAAA
3601 TGCCATTTAC GGCATCCCTT CTGTATGTAA CAAAAAGACA TTCATAATGT
3651 TAGGAAGATG ATAAAAATTC GCTCTTTTAA AGTGCAGCTT ATTATTCTCA
3701 ATTGCTAAAT ACGATTACTC TGCTTTTTTT TTTTCATTTC TTTTGATGTC
3751 ATATGTGAGT ATCTTATAAT TTAGTTCATT TGTTCAAGGT AAAATTTGAA
3801 AAAAAAATTT TACCTGTGC AAAATAGTTT TTTAAAAAT ATACATGTAG
3851 CTCACCTTGA GGTACTGCTA TATAAATATT CACTCACATT ATCACGGAAT
3901 TTATGTATAG TTTCTCTAAT ATAGAAGATA AAATTGGTGT CCTCATAACT
3951 TTAACAAAGA AAACCCCTCAG TCCTATTAT TAATGGGTAG AATTAAATAT
4001 ATAATTTTAT AGCTCAGTTT ACCCAGTATT CATCTGCAAA GCCAGATTGC
4051 TCTCATTGCT TTTATATTTT TAAATGTAG CTTTATGAGA CCTATGATCC
4101 TCATGGAAC TAAATTTTTA TTAATATTC AGGTAACAGT TCTGAATTCA
4151 TGTGATAATG GTGGCATTAT ATATGATTAA ACACCTCAGA ACTTTCTAAT
4201 GTTATCAGGA GTATTTTGAG GGAGATATGA TTATATTGTA TTTTCTCAGA
4251 TAAGAAAAAT GTTTTTTAA AATATTATTT TAATCTGTTT TAAGCATCTC
4301 TTAGATTTAC ATTATAACTA CATAAAGCAG TGAAGCAAAG GCAAAATTAAG
4351 ATAAAGCTAG AAAGCTGAA CATTTTATT CAAATCATA CGAATCGGGG
4401 TCAGTTAAGC CTCAGTATTC TTAGCTTTTG TTGATTTTGG CACTATCTTT
4451 ATATTATTAA ATATATTGTG TGTGTGGATA TTTATATAA AGATGGCTAT
4501 AATTACATAT TTCATTCCCA ATTTGTGTGT GTTGGGGGT ACTTTTAAAG
4551 GTGACTATTG TTTGTACAT CTAATTTTGG GAAACCAAGT CTATAAGACA
4601 TCTGTGATT TCTTAATGTT TTTGTTGTA TGTTTTTCAT AGATATCACT
4651 GTCCTTTATC ATGTTTTGAA GATTGTTTAA AATTCATTTT CCTAAATTA
4701 TGTGCAAGTA ATGTTTTGAG GATATCGGTG TTTTATATTA AACATATTC
4751 CAATTCAAAA AAAAAAATA AAAAACTTAT CGATACCGTC GACCTCGATG
4801 ATGATGATGA TGATGATGAT GTCGAC

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 138 bp to 3089 bp; peptide length: 984
Category: similarity to known protein

1	MDKRLGRPP	SSSEIITEGK	RKKSSSDLSE	IRKMLNAKPE	DVHVQSPLSK
51	FRSSERWTL	QWERSLRN	VLSDHKNKK	HIRGCPVTSR	SSPERIPRVI
101	LTVNLGTEL	RYIRTPPV	EGSLSDTNL	QSEQLSSSD	GSLESYQNLN
151	PHKSCYLSE	GSQSKTVDD	NSAQTAHNN	EKKRKDDGIS	LLISDTQPED
201	LNSGSRGCD	LEQESRNKD	KYSDSKVELT	LISRKTKRRL	RNNLPDSQYC
251	TSLDKSTEQT	KKQEDDSTIS	TEFERPSENY	HQDPKLPEEI	TKPKTKSDFT
301	KLSSLSNSEL	TLNATKSA	AGSTTETVEY	SNSIDIVGIS	SLVEKDENE
351	NTIEKPILRG	HNEGNQSLIS	AEPIVVSSDE	EGPVEHKSSE	ILKLQSKQDR
401	ETTNESTST	ESALLEPLI	TCESVQMSSE	LCPPNPVME	ISSIMPSNEM
451	DLQLDFIFTS	VYIGKIKGAS	KGCVTITKKY	IKIPFQVSLN	EISLLVDTH


```

501 LKRFGWLKSK DDNHSKRSHA ILFFWVSSDY LQEIQTQLEH SVLSQQSKSS
551 EFIFLELHNP VSQREELKLK DIMTEISIIS GELELSYPLS WVQAFPLFQN
601 LSSKESSEFIH YYCVSTCSFP AGVAVAEEMK LKSVSQPSNT DAAKPTYTFL
651 QKSSGCGYSL SITSNPDEEW REVRHTGLVQ KLIVYPPPT KGLGLVTNED
701 LECLEEGEFL NDVIIDFYLK YLILEKASDE LVERSHIFSS FFYKCLTRKE
751 NNLTEDNPNL SMAQRRHKRV RTWTRHINIF NKDYIFVPVN ESSHWYLAVI
801 CFPWLEEAVY EDFPQTVSQO SQAQQSQSDN KTIDNDLRT STLSLSAEDS
851 QSTESNMSVP KKMCKRPCIL ILDSLKAASV RNTVQNLREY LEVEWEVKLK
901 THROFSKTNM VDLCPKVPKQ DNSSDCGVYL LQYVESFFKD PIVNFELPIH
951 LEKWFPRHVI KTKREDIREL ILKLHLQQQK GSSS

```

BLASTP hits

Entry SPAC17A5_7 from database TREMBL:
 "SPAC17A5.07c"; product: "hypothetical protein"; S.pombe
 chromosome I cosmid c17A5. Schizosaccharomyces pombe (fission
 yeast)
 Length = 652
 Score = 275 (96.8 bits), Expect = 1.9e-29, Sum P(3) = 1.9e-29
 Identities = 56/120 (46%), Positives = 78/120 (65%)

Entry S49947 from database PIR:
 SMT4 protein - yeast (Saccharomyces cerevisiae)
 Length = 1034
 Score = 163 (57.4 bits), Expect = 4.6e-16, Sum P(3) = 4.6e-16
 Identities = 46/159 (28%), Positives = 76/159 (47%)

Entry YQG6 CAEEL from database SWISSPROT:
 HYPOTHETICAL 35.7 KD PROTEIN C41C4.6 IN CHROMOSOME II.
 Length = 342
 Score = 162 (57.0 bits), Expect = 6.1e-13, Sum P(3) = 6.1e-13
 Identities = 37/119 (31%), Positives = 62/119 (52%)

Entry AB018340.1 from database TREMBL:
 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens mRNA for
 KIAA0797 protein, partial cds.
 Score = 540, P = 1.9e-50, identities = 120/243, positives = 155/243

Alert BLASTP hits for DKFZphfbr2_16g18, frame 3

TREMBL:ATT16L1.11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII
 project), N = 2, Score = 239, P = 2.1e-18

>TREMBL:ATT16L1.11 gene: "T16L1.110"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone T16L1 (ESSAII project)
 Length = 710

HSPs:

Score = 239 (35.9 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 51/135 (37%), Positives = 78/135 (57%)

Query: 683 IVYPPPTKGLGVNTNEDLECLEEGEFLNDVIIDFYLYLILEKASDELVERSHIFSSFF 742
 +VYP + V +D+E L+ F+ND IIDFY+KYL + S + R H F+ FF
 Sbjct: 176 LVYPQGEPAVV-VRKQDIELLKPRRFINDTIIIDFYIKYL-KNRISPKERGRFHFNCFF 233

Query: 743 YKCLTRKENNLTEDNPNLSMAQRRHKRVRTWTRHINIFNKDYIFVPVNESSHWYLAVICF 802
 + RK NL + P+ + ++RV+ WT+++++F KDYIF+P+N S HW L +IC
 Sbjct: 234 F----RKLANLDKGTPTSCGGREAYQRVQKWKTNVDLFKDYIFIPINCSFHWLSLVIICH 289

Query: 803 PWLEEAVYEDFPQTV 817
 P + + PQ V
 Sbjct: 290 PGELVPSHVENPQRV 304

Score = 70 (10.5 bits), Expect = 2.1e-18, Sum P(2) = 2.1e-18
 Identities = 13/28 (46%), Positives = 15/28 (53%)

Query: 948 PIHLEKWFPFRHVIKTKREDIRELILKLH 975
 P HL WFP KR +I EL+ LH
 Sbjct: 403 PSHLRNWFPAKEASLKRRNILELLYNLH 430

Pedant information for DKFZphfbr2_16g18, frame 3

Report for DKFZphfbr2_16g18.3

[LENGTH] 984
 [MW] 112265.80
 [PI] 6.13
 [HOMOL] TREMBL:AB018340_1 gene: "KIAA0797"; product: "KIAA0797 protein"; Homo sapiens
 mRNA for KIAA0797 protein, partial cds. 8e-53
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YIL031w] 9e-17
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL020c] 4e-06
 [BLOCKS] BL00494C Bacterial luciferase subunits proteins
 [PROSITE] AMIDATION 3
 [PROSITE] MYRISTYL 9
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 30
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] PKC_PHOSPHO_SITE 19
 [PROSITE] ASN_GLYCOSYLATION 12
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 4.47 %

SEQ MDKRLGRPPSSSEIITEGKRKKSSSDLSEIRKMLNAKPEDVHVQSPLSKFRSSERWTL P
 SEG
 PRD cccccccccccccccccccccccccchhhhhhhhhccccccccccccccccccccchh

 SEQ LQWERSLRNKVISLDHKNKKHIRGCPVTSRSSPERIPRVILTNLVGTGLGRKYIRT P PVT
 SEG
 PRD hhhhhhhhhheeecc

 SEQ EGSLSDDTNLQSEQLSSSSDGSLESYQNLNPHKSCYLSERGSQRSKTVDDNSAKQTAHNK
 SEG
 PRD cchhhhhhh

 SEQ EKRRKDDGISLLISDTQPEDLNSGSRGCDHLEQESRNKDVKYSDSKVELTLISRKTKRRL
 SEG
 PRD hhhhhccchhhhhhh

 SEQ RNNLPDSQYCTSLDKSTEQTKQEDDSTISTEFERPSENYHQDKPLPEEITTKPTKSDFT
 SEG
 PRD hccccccccccccccccchhhhhcccccccccccccccccccccccccccccccccccccc

 SEQ KLSSLNSQELTSLNATKSASAGSTTETVEYSNSIDIVGISSLVEKDENEINTEKPIILRG
 SEG
 PRD cccccccccceehhhhhhhcc

 SEQ HNENQSLISAEPVIVSSDEEGPVEHKSEILKLQSKQDRETNENESTSESALLELPLI
 SEG
 PRD cccccccccccccccccccccccccchhhhhhhhhhhhhccccccccchhhhhcccccc

 SEQ TCESVQMSSELCPYNFVMEINISSIMPSNEMDLQLDIFTSVYIGIKIGASKGCVTITKKY
 SEG
 PRD eccccccccccccccccccccccccccccchhhhhhhheeeeeeeeecccccccccccccc

 SEQ IKIPFQVSLNEISLLVDTTHLKRFLWKSDDNHSKRSHAILFFWSSDYLQEIQTQLEH
 SEG
 PRD eeeeeccchhhhhhhhh

 SEQ SVLSQSKSSEFIFLELHNPVSQREELKLDIMTEISIIISGELELSYPLSWVQAFPLFQN
 SEG
 PRD hhhhhccccccccccccccccchhhhhhhhhheeeeecccccccccccccccccccccc

 SEQ LSSKESSFIHYCVSTCSFPAGVAVAEEMKLSVVSQPSNTDAKPTYTFLQKQSSGCYSL
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccccccccc

 SEQ SITSNPDEEWREVHTGLVQKLIVYPPPTKGGLGVTNEDLECEGEFLNDVIIDFYLK
 SEG
 PRD eccccccccccccccccccccccccccccccccccccchhhhhhhhhccccchhhhhhhhh

 SEQ YLILEKASDELVERSHIFSSFFYKCLTRKENNLTEDNPNLSMAQRHRKRVRTWTRHINIF
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhc

 SEQ NKDYIFVPVNESSHWYLAICFPWLEAVYEDFPQTVSQQSQAQSSQSDNKTIDNDLRTT
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhccccchhhhhhhhhcccccccccccccc

 SEQ STLSLSAEDSQSTESNMSVPMCKMCKRPCILILDSLKAASVRNTVQNLREYLEVEWEVKLK
 SEG
 PRD cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhh

 SEQ THRQFSKTNMVDLCPKVPKQDNSSDCGVYLLQYVESFFKDPIVNFELPIHLEKWFPRHVI

```

SEG .....
PRD hhhhhccccccccccccccccccccceeeehhhhhhhccccceccccccccccchh

SEG KTKREDIRELILKLHLQQQKSSS
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhecccc

```

Prosites for DKF2phfbr2_16g18.3

```

PS00001 314->318 ASN_GLYCOSYLATION PDOC00001
PS00001 365->369 ASN_GLYCOSYLATION PDOC00001
PS00001 406->410 ASN_GLYCOSYLATION PDOC00001
PS00001 440->444 ASN_GLYCOSYLATION PDOC00001
PS00001 513->517 ASN_GLYCOSYLATION PDOC00001
PS00001 600->604 ASN_GLYCOSYLATION PDOC00001
PS00001 752->756 ASN_GLYCOSYLATION PDOC00001
PS00001 759->763 ASN_GLYCOSYLATION PDOC00001
PS00001 790->794 ASN_GLYCOSYLATION PDOC00001
PS00001 830->834 ASN_GLYCOSYLATION PDOC00001
PS00001 856->860 ASN_GLYCOSYLATION PDOC00001
PS00001 922->926 ASN_GLYCOSYLATION PDOC00001
PS00004 8->12 CAMP_PHOSPHO_SITE PDOC00004
PS00004 21->25 CAMP_PHOSPHO_SITE PDOC00004
PS00005 54->57 PKC_PHOSPHO_SITE PDOC00005
PS00005 66->69 PKC_PHOSPHO_SITE PDOC00005
PS00005 88->91 PKC_PHOSPHO_SITE PDOC00005
PS00005 158->161 PKC_PHOSPHO_SITE PDOC00005
PS00005 162->165 PKC_PHOSPHO_SITE PDOC00005
PS00005 172->175 PKC_PHOSPHO_SITE PDOC00005
PS00005 233->236 PKC_PHOSPHO_SITE PDOC00005
PS00005 236->239 PKC_PHOSPHO_SITE PDOC00005
PS00005 260->263 PKC_PHOSPHO_SITE PDOC00005
PS00005 291->294 PKC_PHOSPHO_SITE PDOC00005
PS00005 477->480 PKC_PHOSPHO_SITE PDOC00005
PS00005 515->518 PKC_PHOSPHO_SITE PDOC00005
PS00005 562->565 PKC_PHOSPHO_SITE PDOC00005
PS00005 602->605 PKC_PHOSPHO_SITE PDOC00005
PS00005 747->750 PKC_PHOSPHO_SITE PDOC00005
PS00005 874->877 PKC_PHOSPHO_SITE PDOC00005
PS00005 879->882 PKC_PHOSPHO_SITE PDOC00005
PS00005 901->904 PKC_PHOSPHO_SITE PDOC00005
PS00005 962->965 PKC_PHOSPHO_SITE PDOC00005
PS00006 11->15 CK2_PHOSPHO_SITE PDOC00006
PS00006 24->28 CK2_PHOSPHO_SITE PDOC00006
PS00006 91->95 CK2_PHOSPHO_SITE PDOC00006
PS00006 123->127 CK2_PHOSPHO_SITE PDOC00006
PS00006 125->129 CK2_PHOSPHO_SITE PDOC00006
PS00006 137->141 CK2_PHOSPHO_SITE PDOC00006
PS00006 167->171 CK2_PHOSPHO_SITE PDOC00006
PS00006 196->200 CK2_PHOSPHO_SITE PDOC00006
PS00006 225->229 CK2_PHOSPHO_SITE PDOC00006
PS00006 251->255 CK2_PHOSPHO_SITE PDOC00006
PS00006 271->275 CK2_PHOSPHO_SITE PDOC00006
PS00006 295->299 CK2_PHOSPHO_SITE PDOC00006
PS00006 323->327 CK2_PHOSPHO_SITE PDOC00006
PS00006 341->345 CK2_PHOSPHO_SITE PDOC00006
PS00006 377->381 CK2_PHOSPHO_SITE PDOC00006
PS00006 396->400 CK2_PHOSPHO_SITE PDOC00006
PS00006 402->406 CK2_PHOSPHO_SITE PDOC00006
PS00006 408->412 CK2_PHOSPHO_SITE PDOC00006
PS00006 488->492 CK2_PHOSPHO_SITE PDOC00006
PS00006 509->513 CK2_PHOSPHO_SITE PDOC00006
PS00006 536->540 CK2_PHOSPHO_SITE PDOC00006
PS00006 562->566 CK2_PHOSPHO_SITE PDOC00006
PS00006 602->606 CK2_PHOSPHO_SITE PDOC00006
PS00006 638->642 CK2_PHOSPHO_SITE PDOC00006
PS00006 664->668 CK2_PHOSPHO_SITE PDOC00006
PS00006 697->701 CK2_PHOSPHO_SITE PDOC00006
PS00006 747->751 CK2_PHOSPHO_SITE PDOC00006
PS00006 826->830 CK2_PHOSPHO_SITE PDOC00006
PS00006 846->850 CK2_PHOSPHO_SITE PDOC00006
PS00006 962->966 CK2_PHOSPHO_SITE PDOC00006
PS00007 216->223 TYR_PHOSPHO_SITE PDOC00007
PS00008 84->90 MYRISTYL PDOC00008
PS00008 106->112 MYRISTYL PDOC00008
PS00008 141->147 MYRISTYL PDOC00008
PS00008 161->167 MYRISTYL PDOC00008
PS00008 204->210 MYRISTYL PDOC00008
PS00008 468->474 MYRISTYL PDOC00008

```

WO 01/12659

PCT/IB00/01496

PS00008	505->511	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00008	693->699	MYRISTYL	PDOC00008
PS00009	6->10	AMIDATION	PDOC00009
PS00009	18->22	AMIDATION	PDOC00009
PS00009	109->113	AMIDATION	PDOC00009

(No Pfam data available for DKF2phfbr2_16g18.3)

DKF2phfbr2_16112

group: transmembrane protein

DKF2phfbr2_16112 encodes a novel 185 amino acid protein, with strong similarity to PUT2 protein of Fugu rubripes.

The novel protein contains 1 transmembrane region.
PUT 2 is a Fugu rupies protein similar to the neural cell adhesion molecule L1 (L1-CAM) a mitosis-specific chromosome segregation protein (SMC1) and the calcium channel alpha-1 subunit homolog (CCA1).
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

strong similarity to Fugu rubripes PUT2

complete cDNA, complete cds, EST hits,
TRANSMEMBRANE 1

Sequenced by LMU

Locus: /map="873.3/875.1 cR from top of Chri linkage group"

Insert length: 1552 bp

Poly A stretch at pos. 1528, polyadenylation signal at pos. 1506

```
1 GGGGGGGGAC AACTGGGTCT TTTGCGGCTG CAGCGGGCTT GTAGGCGTCC
51 GGCTTTGCTG GCCCAGCAAG CCTGATAAGC ATGAAGCTCT TATCTTTGGT
101 GGCTGTGGTC GGGTGTTTGC TGGTGCCGCC AGCTGAAGCC AACCAAGATT
151 CTGAAGATAT CCGGTGCAAA TGCATCTGTC CACCTTAGAG AACATCAGT
201 GGGCACATTT ACAACCAGAA TGTATCCAGG AAGGACTGTT GTAGCAACTG
251 CCTGCACGTG GTGGAGCCCA TGCCAGTGCC TGGCCATGAC GTGGAGGCCT
301 ACTGCCTGCT GTGCGAGTGC AGGTACGAGG AGCGCAGCAC CACCACCATC
351 AAGGTCATCA TTGTCATCTA CCTGTCCGTG GTGGGTGCCC TGTGTCTCTA
401 CATGGCCTTC CTGATGCTGG TGGACCCCTC GATCCGAAAG CCGGATGCAT
451 AACTGAGGCA ACTGCACAA TGGAGGAGGA ATGAGGATGC TCGCTCTATG
501 GCAGCAGCTG CTGCATCCCT CGGGGGACCC CGAGCAAACA CAGTCCTGGA
551 CCGTGTGGAA GGTGCCCAGC AGCGGTGGAA GCTGCAGGTG CAGGAGCAGC
601 GGAAGACAGT CTTGCATCGG CACAAGATGC TCAGCTAGAT GGGCTGGTGT
651 GGTGTGGTCA AGGCCCAAC ACCATGGCTG CCAGCTTCCA GGCTGGACAA
701 AGCAGGGGGC TACTTCTCCC TTCCCTCGGT TCCAGTCTTC CCTTTAAAG
751 CCTGTGGCAT TTTTCTCCTC TCTCCCTAAC TTTAGAAATG TTGTACTTGG
801 CTATTTTGAT TAGGGAAGAG GGATGTGGTC TCTGATCTCT GTTGTCTTCT
851 TGGGTCTTTG GGGTTGAAGG GAGGGGAAG GCAGGCCAGA AGGGAATGGA
901 GACATTCGAG GCGGCCTCAG GAGTGGATGC GATCTGTCTC TCCTGGCTCC
951 ACTCTTGCCG CCTTCCAGCT CTGAGTCTTG GGAATGTTGT TACCCTTGGA
1001 AGATAAAGCT GGGTCTTCAG GAACTCAGTG TTTGGGAGGA AAGCATGGCC
1051 CAGCATTGAG CATGTGTTCC TTTCTGAGT GGTCTTTATC ACCACCTCCC
1101 TCCCAGCCCC AGCGCCTCAG CCCCAGCCCC AGCTCCAGCC CTGAGGACAG
1151 CTCTGATGGG AGAGCTGGGG CCCCTGAGCC CACTGGGTCT TCAGGGTGCA
1201 CTGGAAGCTG GTGTTGCTG TCCCTGTGCA ACTTCTGCA CTGGGGCATG
1251 GAGTGCCCAT GCATACCTCTG CTGCCGGTCC CCTCACCTGC ACTTGAGGGG
1301 TCTGGGCAGT CCTTCTCTC CCCAGTGTCC ACAGTCACTG AGCCAGACGG
1351 TCGGTTGGAA CATGAGACTC GAGGCTGAGC GTGGATCTGA ACACACAGC
1401 CCCTGTACTT GGGTTGCCCTC TTGTCCTGTA ACTTCGTGT ACCAGTGCAT
1451 GGAGAGAAAA TTTTGTCTCT TTGTCTTAGA GTTGTGTGTA AATCAAGGAA
1501 GCCATCATTA AATTGTTTAA TTTCTCTCAA AAAAAAAAAA AAAAAAATA
1551 TC
```

BLAST Results

Entry HS808349 from database EMBL:
human STS WI-11986.
Score = 1716, P = 5.7e-73, identities = 364/378

Entry HS487355 from database EMBL:
human STS WI-13088.
Score = 1358, P = 1.3e-56, identities = 274/277

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 81 bp to 635 bp; peptide length: 185
Category: similarity to unknown protein

```
1 MKLLSLVAVV GCLLVPPAEA NKSSDIRCK CICPPYRNIS GHIYNQNVSQ
51 KDCSCNCLHV VEPMPVPGHD VEAYCLLCEC RYEERSTTTI KVIIIVYLSV
101 VGALLLYMAF LMLVDPLIRK PDAYTEQLHN EEENEDARSM AAAASLGGP
151 RANTVLERVE GAQQRWKLQV QEQRKTVFDR HKMLS
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16i12, frame 3

TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene,
complete cds; putative protein 1 (PUT1) gene, partial cds;
mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1)
and putative protein 2 (PUT2) genes, partial cds, complete sequence., N
= 1, Score = 655, P = 2.8e-64

TREMBL:CER12C12_5 gene: "R12C12.6"; Caenorhabditis elegans cosmid
R12C12., N = 1, Score = 225, P = 1e-18

>TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu
rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete
cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific
chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and
calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
(PUT2) genes, partial cds, complete sequence.
Length = 187

HSPs:

Score = 655 (98.3 bits), Expect = 2.8e-64, P = 2.8e-64
Identities = 124/163 (76%), Positives = 140/163 (85%)

```
Query: 22 KSSDIRCKCICPPYRNISGHIYNQNVSKDCSCNCLHVVEPMPVPGHDVEAYCLLCECR 81
      KS +D+RCKCICPPYRNISGHIYN+N +QKDC NCLHV+V+PMPVPG+DVEAYCLLCEC+
Sbjct: 31 KSFDVVRCKCICPPYRNISGHIYRNFTQKDC--NCLHVVDPMPPVPGNDVEAYCLLCECK 88

Query: 82 YEERSTTTIKVIIIVYLSVVGALLLYMAFLMLVDPLIRKPDAYTEQLHNEENEDARSM 141
      YEERST TI+V I+I+LSVVGALLLYM FL+LVDPLIRKPD + LHNEE++ED +
Sbjct: 89 YEERSTNTIRVTIIIFLSVVGALLLYMLFLLVDPLIRKPDPLAQLHNEEDSEDIQPM 148

Query: 142 AAAASLGGP-RANTVLERVEGAQQRWKLQVQEQRKTVFDRHKML 184
      + G P R NTVLERVEGAQQRWK QVQEQRKTVFDRHKML
Sbjct: 149 S-----GDPARGNTVLERVEGAQQRWKKQVQEQRKTVFDRHKML 187
```

Pedant information for DKFZphfbr2_16i12, frame 3

Report for DKFZphfbr2_16i12.3

```
[LENGTH] 185
[MW] 20764.29
[pI] 6.21
[HOMOL] TREMBL:AF026198_5 gene: "PUT2"; product: "putative protein 2"; Fugu rubripes
neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1
(PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1)
gene, complete cds; and calcium channel alpha-1 subunit homolog (CCA1) and putative protein 2
(PUT2) genes, partial cds, complete sequence. 3e-68
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 2
[PROSITE] ASN_GLYCOSYLATION 3
[KW] SIGNAL_PEPTIDE 21
```

```

[KW]          TRANSMEMBRANE 1
[KW]          LOW_COMPLEXITY 2.70 %

SEQ  MKLLSLVAVVGCLLVPPAEANKSSEDIRCKCICPPYRNISGHIYNQNVSQKDCCSNCLHV
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  VEPMPVPGHDVEAYCLLCECRYEERSTTTIKVIIIVYLSVVGALLLYMAFLMLVDPLIRK
SEG  .....
PRD  . eccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ  PDAYTEQLHNEEENEDARSMAAAAASLGGPRANTVLERVEGAQQRWKLQVQEQRKTVFDR
SEG  .....xxxxx.....
PRD  cccchhhhhhhhhccccchhhhhhhhhccccccccchhhhhhhchhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  HKMLS
SEG  .....
PRD  hhccc
MEM  .....

```

Prosites for DKF2phfbr2_16i12.3

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	38->42	ASN_GLYCOSYLATION	PDOC00001
PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	148->154	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phfbr2_16i12.3)

DKFZphfbr2_16k22

group: brain derived

DKFZphfbr2_16k22 encodes a novel 108 amino acid protein with very weak similarity to thioredoxin of *Bacillus subtilis*.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to thioredoxin

complete cDNA, complete cds, genomic DNA?
no EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2088 bp
Poly A stretch at pos. 2065, no polyadenylation signal found

```
1 AAAAGGAAGA AGGAAATAAG GATATTTCAA GGGTTACCAA AGTCGAGGAA
51 AACTATTTTA AGAAGAAATC TGAATTTATT GTGCACATAG GTTGTAAATA
101 TAGCATCTTG CATTAATATG TGTTTTCTAG CTTACAAAGT GGATTCATAT
151 AACTATTGTG AACTGACTCT CTACAACTT GCAAGGTTAG CAAGACAAAT
201 GGTATTTTAA GATAACAAAC TGAGACTCAA AAAAGGCAAG TAACTCGTTC
251 TACTTCCCAA AGCCAGAAAG TGGCAAAATA GAAATGGAT CCTGAATCTC
301 CAACACCATG CAAACTAAGA GAGGGAATCC TCTGTAGAGG GAATGGAAGT
351 AAAAAGGCAC AAGTGGTGAT GTCACCTTCT GAACAGAGAT GGAACCTTTC
401 TTCTCTGAG AAAAAGAGA AAAGATAGTT TTAAGTGCCA AAAGAACATG
451 AAGCAATGTG AGGTGAAGAA ACAGAAAAGA CTATGGATGG AATTCCTAGA
501 TGTGAGATAC ACAAAGTTCC ATTCAAAGA GAAATATCTA TAGATAGGCA
551 TAAAGTTACA CACCTGAAGT ACCAACTCTG AACCAGTAAC TCAAGAGATA
601 TTTTGTGTGT CCCACAAGCC ATATGGCTCT GGGGACAAAT TATCTGAAAG
651 TGCCCAATAA GAAAATATT TGAGGAAGGG GAGTGTGTGA GTGAATGAAT
701 TAAAGGACAT CAGAAAGATA CATTGACTGT TCTCCTTCCC AGGAAACAAA
751 GTGGCTAAGT CAAAACAACG GGCAGCTGTG GGATAGCAAA GAAAAAATAA
801 CTTCCAGGCC CAGGTTCTAG TGAAGCTAC TATGGAAGTT AGCCACTCAA
851 CTTAGAACC AGAGGCTTCT TTTCTCCTC CTTCTTATC TTTTCTAGTT
901 TATAGCAAAAT TTATATTGAG CCACCTATTG TTTCTGAATG CTAGTTCCCC
951 TTTAGCATTG CTTTTCTTC ATTCCCTTTG GACTGGCCCA ATGCTTTGGC
1001 CCCTTATCAA AGCATTTTCT AAGAAACAGT CTGACAGCTC TAATTTGCAT
1051 CTGGTTATGC AAGATGTGGT TAAGAACATG GACTCTGGAG GTAAATACAC
1101 CTTGATCCCA ATTCATTCTC TCATTTATTC ATTGAGCAAA TATTTAGTGA
1151 ACATCTAACA TGTGCTAGGC ACTGTTCTAG TTGCTGAGGA TACAGCTTCA
1201 AACAAATATA GGTCTCTGCA AGGATGCCTT CTCTTACCAC TCCTATTGAG
1251 CGTAGTATTG GAAGTCCTGG CCAGGGCAAT CAGGCAAGAA AAAGAAATCA
1301 AGGTCATCCA AATAGGAAGA GAGGAAGTCA AACTATCCCT GTTTACAGAC
1351 AACATGATCC TACATCTAGA AAAAAACCA TTGTCTTAGC CCAAAAGCTT
1401 CTTAGGCTGA TAAACAACCT CAGCAAGTC TTAGGATACA AAATCCATGT
1451 GCAAAAAACA CTAGCATTCT TATACACCAA CAACAGTCAA GCCGAGATCC
1501 AATCAGGAA CAAACTCCTA TTCACAATTG CCACAAAAAC AATAGAACAG
1551 GAAAACAGCT AACTAGGAAG GTGAAAGATC TCTACAAGGA GAACTACAAA
1601 CCACGTGCTA CAGAAATCAG AGATGACACA TATAAATGGA AAAACATTC
1651 ATGATCATGG ATAGGAAGAA TGAATATTAC TGAAATGGCT ATACTGTCCA
1701 AAGCAATTTA TAGATTCAAT GCTATTCTTA GTAAACTACC ATTGAGATTT
1751 TTTACAGAAC TAGAAAAAAA AAAAATATT TTAAGGCTGG GCGCAGTGGC
1801 TCTCACCTGT AATCCCAGCA CTTTGGGAGG CCGAGATGGG TGGATCACGA
1851 GGTGAGGAGA TGGAAAACAT CCTGGCTAAC ATGGTGAAAC CCCGTCTCTA
1901 CTAAAAATAC AAAAAATTAG CCAGGCGTGG TGGTGGGCGC CTGTAATCCC
1951 AGCTGCTCGG GAGGCTGAGG CAGGATAATG GTGTGAACCC GGGAGGCAGA
2001 GCTTGCACTG AGCTGAGATT GCACCACTGC ACTCCAGCCT GAGGGACAGA
2051 GTGAGACTCC ATCTCAAAAA AAAAAAATA AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 832 bp to 1155 bp; peptide length: 108
Category: putative protein

1 MEVSHSTLEP EASFPPFPLS FLVYSKFILS HLFFLNASSP LAFLFLHSLW
51 TGPMLWPLIK AFSKKQSDSS NLHLVMQDVV KNMDSGGKYT LIPIHSLIYS
101 FSKYLVNI

BLASTP hits

Entry B37192 from database PIR:
thioredoxin - Bacillus subtilis Score = 71 (25.0 bits), Expect = 0.040,
P = 0.039
Identities = 16/49 (32%), Positives = 30/49 (61%)

Alert BLASTP hits for DKFZphfbr2_16k22, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_16k22, frame 1

Report for DKFZphfbr2_16k22.1

[LENGTH] 108
[MW] 12281.47
[pI] 8.06
[PROSITE] MYRISTYL 1
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MEVSHSTLEPEASFPPFPLSFLVYSKFILSHLFFLNASSPLAFLFLHSLWTGPMLWPLIK
PRD cccccccccccccccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhccccchhhhh

SEQ AFSKKQSDSSNLHLVMQDVVKNMDSGGKYTLIPIHSLIYSFSKYL VNI
PRD hhhccccccccceehhhhhccccccccceeeccceeecccccccc

Prosite for DKFZphfbr2_16k22.1

PS00001	36->40	ASN_GLYCOSYLATION	PDOC00001
PS00004	64->68	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00006	6->10	CK2_PHOSPHO_SITE	PDOC00006
PS00008	86->92	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_16k22.1)

DKF2phfbr2_16112

group: transmembrane protein

DKF2phfbr2_16112 encodes a novel 267 amino acid protein with similarity to gallus gallus putative transmembrane protein E3-16

The novel protein contains one putative transmembrane domain. In chicken, E3-16 is expressed specifically in the inner ear.

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neurons involved in perception of hearing.

similarity to gallus putative transmembrane protein E3-16

complete cDNA, complete cds, EST hits
potential start at Bp 73 matches kozak consensus PyCCataG
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 2042 bp

Poly A stretch at pos. 2024, polyadenylation signal at pos. 2003

```
1 GGGGGCGGCG GAGGCAGAGA CCGAGGCTGC ACCGGCAGAG GCTGCGGGGC
51 GGACGCGCGG GCCGCGCGAG CCATGGTGAA GATTAGCTTC CAGCCCGCCG
101 TGGCTGGCAT CAAGGGCGAC AAGGCTGACA AGGCGTCGGC GTCGGCCCTT
151 GCGCCGGCCT CGGCCACCGA GATCCTGCTG ACGCCGGCTA GGGAGGAGCA
201 GCCCCACAAA CATCGATCCA AGAGGGGGGG CTAGTGGGCG GCGGTGTGCT
251 ACCTGTCGAT GGGCATGGTC GTGCTGCTCA TGGGCCTCGT GTTCGCTCTT
301 GTCTACATCT ACAGATACTT CTTCTTGGC CAGCTGGCCC GAGATAACTT
351 CTTCCGCTGT GGTGTGCTGT ATGAGGACTC CCTGTCTCTC CAGGTCCGGA
401 CTCAGATGGA GCTGGAAGAG GATGTGAAAA TCTACCTCGA CGAGAACTAC
451 GAGCGCATCA ACGTGCTGTG GCGCCAGTTT GCGCGCGGTG ACCCTGCAGA
501 CATCATCCAT GACTTCCAGC GGGGTCTGAC TGCCTACCAT GATATCTCCC
551 TGGACAAGTG CTATGTCATC GAACTCAACA CCACCATGTG GCTGCCCCCT
601 CGCAACTTCT GGGAGCTCCT CATGAACGTG AAGAGGGGGA CCTACCTGCC
651 GCAGACGTAC ATCATCCAGG AGGAGATGGT GGTACCGGAG CATGTCAGTG
701 ACAAGGAGGC CTGGGGTCC TTCATCTACC ACCTGTGCAA CGGGAAAGAC
751 ACCTACCGGC TCCGGCGCCG GGCACGCGCG AGGCGGATCA ACAAGCGTGG
801 GGCCAGAAC TGCAATGCCA TCCGCCACTT CGAGAACACC TTCGTGGTGG
851 AGACGCTCAT CTGCGGGGTG GTGTGAGGCC CTCCTCCCCC AGAACCCCTT
901 GCGGTGTCTC TCTTTCTTC TTTCCGGCTG CTCTCTGCCC CTCCTCTTCT
951 CCGCTGCTTA GCTTGTACTT TGGACGGGTT TCTATAGAGG TGACATGTCT
1001 CTCCATTCCCT CTCCAACCCCT GCCCACCTCC CTGTACCAGA GCTGTGATCT
1051 CTCGGTGGGG GGCCCATCTC TGCTGACCTG GGTGTGGCGG AGGGAGAGGC
1101 GATGCTGCAA AGTGTCTTCT GTGTCCCACT GTCTTGAAGC TGGGCTGCC
1151 AAAGCCTGGG CCCACAGCTG CACCGGCAGC CCAAGGGGAA GGACCGGTTG
1201 GGGGAGCCGG GCATGTGAGG CCCTGGGCAA GGGGATGGGG CTGTGGGGGC
1251 GGGGCGGCAT GGGCTTCAGA AGTATCTGCA CAATTAGAAA AGTCCTCAGA
1301 AGCTTTTCTT TGGAGGGTAC ACTTCTTCTA CTGTCCCTAT TCCTAGACCT
1351 GGGGCTTGAG CTGAGGATGG GACGATGTGC CCAGGGAGGG ACCCACCAGA
1401 GCACAAGAGA AGGTGGCTAC CTGGGGGTGT CCCAGGGAAT CTGTCACTGC
1451 CTTACGCCCC CCAGCAGGAG CTTGGAGTTT GGGGAGTGGG GATGAGTCCG
1501 TCAAGCACAA CTGTTCTCTG AGTGGAAACA AAGAAGCAAG GAGCTAGGAC
1551 CCCCAGTCTT GCCCCCCAGG AGCACAAACA GGGTCCCTCT AGTCAAGGCA
1601 GTGGGATGGG CCGCTGAGGA ACGGGGCAGG CAAGGTCACT GCTCAGTCAC
1651 GTCCACGGGG GACGAGCCGT GGGTCTGCTT GAGTAGGTGG AGCTCATTCG
1701 TTTCTCCAAG CTTGGAAGTG TTTTGAAGA TAACACAGAG GGAAAGGGAG
1751 AGCCACCTGG TACTTGTCCA CCCTGCCTCC TCTGTTCTGA AATTCCATCC
1801 CCCTCAGCTT AGGGGAATGC ACCTTTTCTC CTTTCTCTCT CACTTTTGCA
1851 TGTTTTTACT GATCATTCGA TATGCTAACC GTTCTCAGCC CTGAGCCTTG
1901 GAGAGGAGGG CTGTAACGCC TTCAGTCAGT CTCTGGGGAT GAAACTCTTA
1951 AATGCTTTGT ATATTTTCTC AATTAGATCT CTTTCTCAGAA GTGCTATAG
2001 AACAATAAAA ATCTTTTACT TCTGAAAAA AAAAAAAAAA AA
```

BLAST Results

No BLAST result

Medline entries

96325063:
Isolation of markers for chondro-osteogenic differentiation
using cDNA library subtraction. Molecular cloning and
characterization of a gene belonging to a novel multigene
family of integral membrane proteins.

Peptide information for frame 1

ORF from 73 bp to 873 bp; peptide length: 267
Category: similarity to known protein

```

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK
51 RGGSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY
101 EDLSLSSQVRT QMELEEDVKI YLDENYERIN VVPVQFGGDD PADIHDFQR
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFV ELLMNVKRGY YLPQTYIIQE
201 EMVTEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI
251 RHFENTFVVE TLICGVV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_16112, frame 1

SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16)., N = 1, Score = 573, P = 1.4e-55

SWISSNEW:ITMB_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N =
1, Score = 559, P = 4.2e-54

SWISSNEW:ITMA_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1,
Score = 452, P = 9.1e-43

>SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN
E3-16).
Length = 262

HSPs:

Score = 573 (86.0 bits), Expect = 1.4e-55, P = 1.4e-55
Identities = 118/264 (44%), Positives = 175/264 (66%)

```

Query:      1 MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY 60
            MVK+SF  A+A  + A+K  ++      ++L+ P  + + P+      G  C+
Sbjct:      1 MVKVSFNSALA--HKEAANKEEENS-----QVLILPP-DAKEPEDVVVPAGHKRAWCW 50

Query:      61 -LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL- ----SQVRTQM- 112
            +  G+  +L G++  Y+Y+YF  Q  + CG+ Y ED LS  +Q+++
Sbjct:      51 CMCFLAFLAGVILGGAYLYKYFAFQQ---GGVYFCGIKYIEDGLSLPESGAQLKSARY 107

Query:      113 -ELEEDVKIYLDENYERINVPVQFGGDPADIHDFQRLTAYHDISLDKCYVIELNTT 171
            +E++++I  +E+ E I+VPVP+F  DPADI+HDF R LTAY D+SLDKCYVI LNT+
Sbjct:      108 HTIEQNIQILEEEDVEFISVPVPEFADSDPADIVHDFHRLTAYLDLSLDKCYVIPLNTS 167

Query:      172 IVLPPRNFWEELLMNVKRGTYLPQTYIIQEEMVTEHVSDKEALGSFIYHLCNGKDTYRLR 231
            +V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+
Sbjct:      168 VMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVVDQLGFFIYRLCRGKETYLQ 227

Query:      232 RRATRRRINKRGAKNCNAIRHFENTFVVETLIC 264
            R+  + I KR A NC  IRHFEN F +ETLIC
Sbjct:      228 RKEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

```

Pedant information for DKFZphfbr2_16112, frame 1

Report for DKFZphfbr2_16112.1

[LENGTH] 267
[MW] 30223.94

```

[pI]          8.16
[HOMOL]       SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
le-49
[PROSITE]     PRENYLATION      1
[PROSITE]     MYRISTYL        5
[PROSITE]     CAMP_PHOSPHO_SITE  2
[PROSITE]     CK2_PHOSPHO_SITE  3
[PROSITE]     TYR_PHOSPHO_SITE  1
[PROSITE]     PKC_PHOSPHO_SITE  4
[PROSITE]     ASN_GLYCOSYLATION 1
[KW]          TRANSMEMBRANE 1
[KW]          LOW_COMPLEXITY  15.36 %

```

```

SEQ  MVKISFQPAVAGIKGDKADKASASAPAPASATEILLTPAREEQPPQHRSKRGGSVGGVCY
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  cccccccccchhhhhhhhhhhhhhhccccccccccccccccccccccccccccchh
MEM  .....MMMMMMMMM.....

```

```

SEQ  LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLVEDSLSSQVRTQMELEEDVKI
SEG  .....XXXXXXXXX.....
PRD  hhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhccccccccccccchhhhhhhhhhhhh
MEM  MMMMMMMMMMMMMMMMM.....

```

```

SEQ  YLDENYERINVPVPQFGGDPADIIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNFV
SEG  .....
PRD  hhccccccccccccccccchhhhhhhhhhhhhccccccccccccccccccccchh
MEM  .....

```

```

SEQ  ELLMNVKRGTYLPQTYIIQEEMVVTEHVSDEALGSFIYHLCNGKDTYRLRRRATRRRIN
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhccccccccccccccccchhhhhhhccccchhhhhheccccchhhhhhhhhhhhh
MEM  .....

```

```

SEQ  KRGAKNCNAIRHFENTFVVETLICGVV
SEG  xx.....
PRD  hhhccccccccccccchhhhhhecccc
MEM  .....

```

Prosites for DKFZphfbr2_16112.1

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	53->59	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphfbr2_16112.1)

DKF2phfbr2_22f21

group: brain derived

DKF2phfbr2_22f21 encodes a novel 567 amino acid protein with weak similarity to C. elegans cosmid C18C4.5

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to C.elegans C18C4.5

EST HSAA6531/HSAA5273/ defines splice variant, or unspliced cDNA additional ~180 Bp at position 250

Sequenced by AGOWA

Locus: /map="311.4 cR from top of Chr14 linkage group"

Insert length: 1910 bp

Poly A stretch at pos. 1887, polyadenylation signal at pos. 1867

```
1 TGGGCCCTTA GCAACGGCCT GCGACGGTT TCCTTGCTGC TGCAGCCCCC
51 GTCGGCTCCT CTTTCCAGT CCTCCACTGC CGGGGCTGGG CCCGGCCGGG
101 GGAAGGACCG AAGGGGATAC AGCGTGTCCT TCGCGCGCGT GCAAGAGGAC
151 TAAGCATGGA TGGCAGCCGG AGAGTCAGAG CAACCTCTGT CCTCCCAGA
201 TATGGTCCAC CGTGCCCTATT TAAAGGACAC TTGAGCACC AAGTAATGC
251 TGCAAGTAGAC TGCTCGGTTT CAGTAAGCAT GAGTACCAGC ATAAAGTATG
301 CAGACCAACA ACGAAGAGAG AAACCTCAAAA AGGAATTAGC ACAATGTGAA
351 AAAGAGTTCA AATTAACATA AACTGCAATG CGAGCCAATT ATAAAAATAA
401 TTCCAAGTCA CTTTTAATA CCTTACAAGA GCCCTCAGGC GAACCGCAAA
451 TTGAGGATGA CATGTTAAAA GAAGAAATGA ATGGATTTTC ATCCTTTGCA
501 AGGTCACCTAG TACCTCTTTC AGAGAGACTA CACCTAAGTC TACATAAATC
551 CAGTAAAGTC ATCACAATG GTCCTGAGAA GAACCTCAGT TCCTCCCCGT
601 CCAGTGTGGA TTATGCAGCC TCCGGGCCCC GGAAACTGAG CTCTGGAGCC
651 CTGTATGGCA GAAGGCCCCAG AAGCACATTC CCAAAATCCC ACCGGTTTCA
701 GTTAGTCATT TCGAAAGCAC CCAGTGGGGA TCTTTGGAT AAACATTCTG
751 AACTCTTTTC TAACAAACAA TTGCCATTCA CTCCTCGCAC TTTAAAAACA
801 GAAGCAAAAT CTTTCCTGTC ACAGTATCGC TATTATACAC CTGCCAAAAG
851 AAAAAAGGAT TTTACAGATC AACGGATAGA AGCTGAAACC CAGACTGAAT
901 TAAGCTTTAA ATCTGAGTTG GGGACAGCTG AGACTAAAAA CATGACAGAT
951 TCAGAAATGA ACATAAAGCA GGCATCTAAT TGTGTGACAT ATGATGCCAA
1001 AGAAAAATA GCTCCTTTAC CTTTAGAAGG GCATGACTCA ACATGGGATG
1051 AGATTAAGGA TGATGCTCTT CAGCATTCCT CACCAAGGGC AATGTGTGAG
1101 TATTCCTGTA AGCCCTCTTC AACTCGTAAA ATCTACTCTG ATGAAGAAGA
1151 ACTGTGTGAT CTGAGTTTCA TTGAAGATGT AACAGATGAA ATTTTGAAAC
1201 TTGTTTATT TTTAAACAGG TTTTAGAAC GACTGTTTCA GCGACATATA
1251 AAACAAATA AACATTTGGA GGGGAAAAA ATGCCGCCAC TGCTGCATGT
1301 CCTGAAAGTA GACTTAGGCT GCACATCGGA GGAAACTCG GTAAAGCAAA
1351 ATGATGTTGA TATGTTGAAT GTATTGATT TTGAAAGGC TGGGAATTCA
1401 GAACCAATA AATTAATAA TGAAGTGAA GTAACAATC AGCAGGAACG
1451 TCAACAATAC CAAAGGCTT TGGATATGTT ATTGTCGCA CCAAGGATG
1501 AGAACGAGAT ATTCCTTCA CCAACTGAAT TTTTCATGCC TATTTATAAA
1551 TCAAAGCATT CAGAAGGGT TATAATTCAA CAGGTGAATG ATGAACAAA
1601 TCTTGAAACT TCAACTTTGG ATGAAATCA TCCAAGTATT TCAGACAGTT
1651 TAACAGATCG GGAACTTCT GTGAATGTCA TTGAAGGTGA TAGTGACCCT
1701 GAAAGGTTG AGATTTCAAA TGGATTATGT GGTCTTACA CATCACCTC
1751 CCAATCTGTT CAGTCTCTCA GTGTCAAAGG CGACAATAAT CATGACATGG
1801 AGTTATCAAC TCTTAAATC ATGGAATGA GCATTGAGGA CTGCCCTTTG
1851 GATGTTTAA CTTCATTAAT AAATACCTCA AATGGCCAGT AAAAAAAAAA
1901 AAAAAAAAAA
```

BLAST Results

Entry HS477360 from database EMBL:

human STS WI-14643.

Length = 418

Minus Strand HSPs:

Score = 1850 (277.6 bits), Expect = 2.5e-77, P = 2.5e-77

Identities = 392/405 (96%), Positives = 392/405 (96%), Strand = Minus / Plus

.....

Peptide information for frame 3

1	MDGSRVRVRAT	SVLPRYGPPC	LFKGHLSTKS	NAAVDCSVPV	SMSTSIKYAD
51	QORREKKLKE	LQACEKEFL	TKTAMRANLY	NNSKSLFNTL	QEPGSGEPQIE
101	DDMLKEEMNG	FSQFARSIVL	SSERLHLSHY	KSKSVLPTNG	ENKSSSSPQEE
151	VDYAASGPRK	LSGGLAGKFR	PRSTFPNSHR	FQLVKISFQD	GDLLDKHSLE
201	FSNKGLPFTP	RTLTKEARSF	LSQRYTYPTA	KRQVDFDQOR	IEAETQTELS
251	FSKSELGTAEI	KNMTDSEMMI	KQASNCVYD	AKKEIAPLEL	EKGHDSTWDEI
301	KDDALQHSSP	RAMCQYSKLP	STRKTIYSDE	EELLYLSFIE	DVTDEILKIG
351	LFSNRFLRLR	FERHIKQKNH	LEGEKMRHLH	HVLKVDLGCT	SEENSVKQND
401	VDMLNVDFDE	KAGNESPNNL	KNESEVTIQQ	EQYQKQALD	MLLSAPKDEN
451	ETFPSPTEFF	MPYIKXSHSE	GVTIQQVND	TNETSTLESD	NHPSISDLSL
501	DRETSVMNVEI	EDSDPEKVEI	SNGLCGLNTS	PSQSVQFSSV	KGDNNHDMEL
551	STLKMNPMST	EGDCLPDV			

Entry CEC18C4_3 from database TREMBL:
"C18C4.5"; *Caenorhabditis elegans* cosmid C18C4.
Length = 1091
Score = 98 (34.5 bits), Expect = 0.29, P = 0.25
Identities = 105/470 (22%), Positives = 192/470 (40%)

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 22f21, frame 3

[LENGTH]	567	
[MW]	64120.02	
[pI]	5.68	
[PROSITE]	AMINATION	1
[PROSITE]	MYRISTYL	3
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	16
[PROSITE]	PKC_PHOSPHO_SITE	18
[PROSITE]	ASN_GLYCOSYLATION	4
[KW]	All Alpha	
[KW]	LOW_COMPLEXITY	1.23 %

153

```

SEQ      KDDALQHSSPRAMCQYSLKPPSTRKIYSDDEEALLYLSFIEDVTOEILKGLGFSNRFLERL
SEG
PRD      cccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhh

SEQ      FERHIKQNKHLEGEKMRHLLHLVLKVDLGCTSEENSVKQNDVDMLNVDFEKAGNSEPNKL
SEG
PRD      hhhhhhhhhhhccchhhhhhhhhccccccccccccccccccccccceeecccccccccccccc

SEQ      KNESEVTIQQERQQYKALDMLLSAPKDENEIFPSPTeffFMPiYKSKHSEGVIIQQVND
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccceeecccccc

SEQ      TNLETSTLDENHPSISDSLTDRETSVNVNIEGSDSPEKVEISNGLCLGNTSPSQSVQFSSV
SEG
PRD      cccccccccccccccccccccccccccccccccceccccccccccccccccccccccccceeecccc

SEQ      KGDNNHDMELSTLKIMEMSIEDCPLDV
SEG
PRD      cccccchhhhhhhhhhhhhhhhhcccccc

```

Prosites for DKFZphfbr2_22f21.3

PS000001	31->385	ASN_GLYCOSYLATION	PDOC000001
PS000001	143->147	ASN_GLYCOSYLATION	PDOC000001
PS000001	262->266	ASN_GLYCOSYLATION	PDOC000001
PS000001	422->426	ASN_GLYCOSYLATION	PDOC000001
PS000004	159->163	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	4->7	PKC_PHOSPHO_SITE	PDOC000005
PS000005	27->30	PKC_PHOSPHO_SITE	PDOC000005
PS000005	45->48	PKC_PHOSPHO_SITE	PDOC000005
PS000005	122->125	PKC_PHOSPHO_SITE	PDOC000005
PS000005	132->135	PKC_PHOSPHO_SITE	PDOC000005
PS000005	178->181	PKC_PHOSPHO_SITE	PDOC000005
PS000005	202->205	PKC_PHOSPHO_SITE	PDOC000005
PS000005	209->212	PKC_PHOSPHO_SITE	PDOC000005
PS000005	212->215	PKC_PHOSPHO_SITE	PDOC000005
PS000005	250->253	PKC_PHOSPHO_SITE	PDOC000005
PS000005	309->312	PKC_PHOSPHO_SITE	PDOC000005
PS000005	317->320	PKC_PHOSPHO_SITE	PDOC000005
PS000005	322->325	PKC_PHOSPHO_SITE	PDOC000005
PS000005	353->356	PKC_PHOSPHO_SITE	PDOC000005
PS000005	395->398	PKC_PHOSPHO_SITE	PDOC000005
PS000005	500->503	PKC_PHOSPHO_SITE	PDOC000005
PS000005	539->542	PKC_PHOSPHO_SITE	PDOC000005
PS000005	552->555	PKC_PHOSPHO_SITE	PDOC000005
PS000006	89->93	CK2_PHOSPHO_SITE	PDOC000006
PS000006	149->153	CK2_PHOSPHO_SITE	PDOC000006
PS000006	245->249	CK2_PHOSPHO_SITE	PDOC000006
PS000006	264->268	CK2_PHOSPHO_SITE	PDOC000006
PS000006	295->299	CK2_PHOSPHO_SITE	PDOC000006
PS000006	328->332	CK2_PHOSPHO_SITE	PDOC000006
PS000006	337->341	CK2_PHOSPHO_SITE	PDOC000006
PS000006	390->394	CK2_PHOSPHO_SITE	PDOC000006
PS000006	455->459	CK2_PHOSPHO_SITE	PDOC000006
PS000006	481->485	CK2_PHOSPHO_SITE	PDOC000006
PS000006	486->490	CK2_PHOSPHO_SITE	PDOC000006
PS000006	494->498	CK2_PHOSPHO_SITE	PDOC000006
PS000006	498->502	CK2_PHOSPHO_SITE	PDOC000006
PS000006	500->504	CK2_PHOSPHO_SITE	PDOC000006
PS000006	513->517	CK2_PHOSPHO_SITE	PDOC000006
PS000006	559->563	CK2_PHOSPHO_SITE	PDOC000006
PS000008	164->170	MYRISTYL	PDOC000008
PS000008	256->262	MYRISTYL	PDOC000008
PS000008	350->356	MYRISTYL	PDOC000008
PS000009	167->171	AMIDATION	PDOC000009

(No Pfam data available for DKFZphfbr2_22f21.3)

DKFZphfbr2_22h13

group: transmembrane protein

DKFZphfbr2_22h13 encodes a novel 520 amino acid protein, with similarity to *Drosophila melanogaster* EG:39E1.3.

The protein contains an ATP/GTP A Prosite pattern (P-loop). This loop interacts with one of the phosphate groups of a A or G nucleotide. It is found in numerous ATP- or GTP-binding proteins, such as ATP synthase alpha and beta subunits, Myosin heavy chains, Kinesin heavy chains and kinesin-like proteins, Dynamins and dynamin-like proteins, several kinases, DNA and RNA helicases, GTP-binding elongation factors and the Ras family of GTP-binding proteins. Additionally, the novel protein contains one putative transmembran domain.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

AC004780_1, differences to predicted genmodel

membrane regions: 1

AC004780_1, differences to predicted genmodel

complete cDNA, complete cds, EST hits
on genomic level encoded by AC004780,
differences to predicted genmodel!
TRANSMEMBRANE 1

Sequenced by AGOWA

Locus: unknown

Insert length: 2292 bp

Poly A stretch at pos. 2272, polyadenylation signal at pos. 2255

```

1 GGGGGAGGGA ACTGATCTCA GCTCGGGCCC GCGTTACATC CTCCTCCTCT
51 TCTTCCTTCG GCCCAGCTTT CCTTAGGGGC TGCAACCCCG ACGCCGAGGC
101 CGGTTTCGGA GTGGGGAGTG CCCATTITCT CTCCTTCCCA CGTTCCTGGC
151 CCCCAGACGC CATTTGACAG CGGGTGGCTT GGGTCAGCCT CCCC GCCCCC
201 ACCCGACTCC CGTCACGGGA GAGCGCACAC CGCGCCCCGA GAACCAATCA
251 GCAGCCGCGT TAGGTAACCA TGTCTGAGTC TGGACACAGT CAGCCTGGAC
301 TCTATGGGAT AGAGCGGCGG CGACGGTGA AGGAGCCTGG CTCTGGTGGC
351 CCCCAGAATC TCTCTGGGCC TGCTGGTCTGG GAGAGGGACT ACATTGCACC
401 ATGGGAAGA GAGAGAAGGG ATGCCAGCGA AGAGACAAGC ACTTCCTGCA
451 TGCAGAAAAA CCCCATCATC CTCTCAAAAC CTCCAGCAGA GCGGTCAAAA
501 CAGCCACCAC CTCCAACAGC CCCTGCTGCC CCGCCTGCCT CAGCCCTCTT
551 GGAGAAAGCCC ATCGTTCTCA TGAAGCCACG GGAGGAGGGG AAGGGGCTTG
601 TGGCCGTGAC AGGTGCCTCT ACCCTGAGG GCACCGCCCC ACCACCCCTT
651 GCAGCCCTTG CGCCACCCAA GGGGAGAAG GAGGGGAGA GACCCACACA
701 GCCTGTGTAC CAGATCCAGA ACCGGGGCAT GGGCACTGCC GCACCAGCAG
751 CCATGGACCC TGTCGTGGGT CAGGCCAAAC TACTGCCCCC AGAGCGCATG
801 AAGCACAGCA TCAAGTTGGT GGATGACCAG ATGAATTGGT GTGACAGTGC
851 CATCGAGTAC CTGTTGGATC AGACTGATGT GTTGGTGGTT GGTGTCTCTG
901 GCCTCCAGGG GACAGGCAAG TCCATGGTCA TGTCTATTGT GTGAGCCAAC
951 ACTCCAGAGG AGGACCAAGG GACTTATGTT TTCCGGGCCC AGAGCGCTGA
1001 AATGAAGGAA CGAGGGGGCA ACCAGACCA GGCATCGAC TTCTTTATTA
1051 CCCAAGAAGC GATTGTTTTT CTGGACACAC AGCCCATCCT GAGCCCTTCT
1101 ATCCTAGACC ATCTCATCAA TAATGACCGC AAAGTGCCTC CAGAGTACAA
1151 CCTTCCCCAC ACTTACGTTG AAATGCAGTC ACTCCAGATT GCTGCTTCC
1201 TTTTCACGGT CTGCCATGTG GTGATTGTTG TCCAGGACTG GTTCACAGAC
1251 CTCAGTCTCT ACAGGTTCTT GCAGACAGCA GAGATGGTGA AGCCCTCCAC
1301 CCCATCCCCC AGCCACGAGT CCAGCAGCTC ATCGGGCTCC GATGAAGGCA
1351 CCGAGTACTA CCCCACCTA GTCTTCTTGC AGAACAAAGC TCGCCGAGAG
1401 GACTTCTGTC CTCGGAAGCT GCGGCAGATG CACCTGATGA TTGACCAGCT
1451 CATGGCCACC TCCCACCTGC GTTACAAGGG AACTCTGTCC ATGTTACAAT
1501 GCAATGTCTT CCGGGGGCTT CCACCTGACT TCCTGGACTC TGAGGTCAAC
1551 TTATTCTCTG TACCCCTTAT GGACAGTGAA GCAGAGAGTG AAAACCCACC
1601 AAGAGCAGGA CTTGGTTCCA GCCACTCTT CTCCTGCTG CTTGGGTATC
1651 GTGCCACACC CAGTTTCCAG TCCTTGGTGA GCAAGTCCG GAGCCAAGTG
1701 ATGTCCATGG CCGGCCACA GCTGTACAC ACGATCCTCA CCGAGAAGAA
1751 CTGGTTCCAC TACGCTGCCC GGATCTGGGA TGGGGTGAGA AAGTCTCTG
1801 CTCTGGCAGA GTACAGCCGC CTGCTGGCCT GAGGCCAAGG AGAGGAATGT
1851 CATGCAGGGG ACCTCCTGGG TCCGAGTGT ACTGCGAGGG AGCACAGATG
1901 TCCATCCCCC GCTGGGGTGG AGAGCGGCAG CAGGCCTGAT GGATGAGGGA
1951 TCGTGGCTTC CCGGCCAGA GACATGAGGT GTCCAGGGCC AGGCCCCCA

```



```

2001 CCCTCAGTTG GGGCTGTTCC GGGGGTGA CTGTGAGCGATC CCACCCCAAA
2051 CCTGAGATGG GGTAGCCCGT CCTGTGTCCT CCACAGGGAC AAGCAGTGGG
2101 AGGAGTCTGA ATGGTCACCA GGAAGCCCGG GCTCCATCTT GACCTCCTTT
2151 TTCAGGGACA GGAGCAACAG GCCCTCTTCT CCGTACTCTA AGCCCTTCCC
2201 TGTAAGGTGA GGCAGGGTCT GGAGAGCTCT TTATTGGAAC AGATCTGGTG
2251 GTTCAAAATA ACACAGTCAT GCAAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry AC004780 from database EMBL:
Homo sapiens chromosome 19, cosmid F17127, complete sequence.
Score = 2616, P = 0.0e+00, identities = 524/525
15 exons Bp 8031-31789

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 270 bp to 1829 bp: peptide length: 520
Category: similarity to unknown protein
Prosite motifs: ATP_GTP_A (211-219)

```

1 MSEGHSQPG LYGIERRRRW KEPGSGGPON LSGPGGRERD YIAPWERERR
51 DASEETSTSV MQKTPIILSK PPAERSKQPP PPTAPAAPPA PAPLEKPIVL
101 MKPREEGKGP VAVTGASTPE GTAPPPPAAP APPKGEKEGQ RPTQPVYQIQ
151 NRGMGTAAPA AMDPVVGQAK LLPPERMKHS IKLVDDQMNW CDSAIEYLLD
201 QTDVLVVGVL GLQGTGKSMV MSLLSANTPE EDQRTYVFRA QSAEMKERGG
251 NQTSIGIDFFI TQERIVFLDT QPILSPSILD HLINNDRLKP PEYNLPHTYV
301 EMQSLQIAAF LFTVCHVVIV VQDWFTDLSL YRFLQTAEMV KPSTPSPSHE
351 SSSSSGSDEG TEYYPHLVFL QNKARREDFC PRKLQRMHLM IDQLMAHSHL
401 RYKGTLSMLQ CNVFPGLPFD FLDSEVNLFV VPFMDSEAES ENPPRAGPGS
451 SPLFSLLPYR RGHFSPFQSLV SKLRSQVMSM ARPQLSHTIL TEKNWFHYAA
501 RIWDGVRKSS ALAEYSRLLA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_22h13, frame 3

TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19,
cosmid F17127, complete sequence., N = 2, Score = 1264, P = 1.3e-231

TREMBL:CEY54E2A_1 gene: "Y54E2A.2"; Caenorhabditis elegans cosmid
Y54E2A, N = 2, Score = 219, P = 1.4e-15

>TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence.
Length = 528

HSPs:

Score = 1264 (189.6 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
Identities = 254/302 (84%), Positives = 264/302 (87%)

Query: 46 ERERRDASEETSTSVMQKTPIILSKPPAERSKQPPPTAPAAPPAAPLEKPIVIMKPRE 105
E+ER D+ + S +Q+T + R + P + A APLEKPIVIMKPRE
Sbjct: 39 EKER-DSDSDFSP--LQTEGCRDRKHFRAENPHPLKTSSRA-APLEKPIVIMKPRE 94

Query: 106 EGKGPVAVTGASTPEGTAPPPPAAPPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV 165
EGKGPVAVTGASTPEGTAPPPPAAPPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV
Sbjct: 95 EGKGPVAVTGASTPEGTAPPPPAAPPPKGEKEGQRPTQPVYQIQNRGMGTAAAPAMDPV 154

Query: 166 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 225
VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS
Sbjct: 155 VGQAKLLPPERMKHSIKLVDDQMNWCDSAIEYLLDQTDVLVVGVLGLQGTGKSMVMSLLS 214

```

Query:      226 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTPILSPSILDHLINN 285
Sbjct:      215 ANTPEEDQRTYVFRAQSAEMKERGGNQTSGIDFFITQERIVFLDTPILSPSILDHLINN 274

Query:      286 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRLQTAEMVKPSTP 345
          DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYR      K ++
Sbjct:      275 DRKLPPEYNLPHTYVEMQSLQIAAFLFTVCHVVIVVQDWFTDLSLYRLWDLGCKCKSNSH 334

Query:      346 SP 347
          SP
Sbjct:      335 SP 336

Score = 993 (149.0 bits), Expect = 1.3e-231, Sum P(2) = 1.3e-231
Identities = 189/189 (100%), Positives = 189/189 (100%)

Query:      332 RFLQTAEMVKPSTPSPSHSSSSSSGSGDEGTEYYPHLVFLQNKARREDFCPKRLRQMHLM 391
          RFLQTAEMVKPSTPSPSHSSSSSSGSGDEGTEYYPHLVFLQNKARREDFCPKRLRQMHLM
Sbjct:      340 RFLQTAEMVKPSTPSPSHSSSSSSGSGDEGTEYYPHLVFLQNKARREDFCPKRLRQMHLM 399

Query:      392 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 451
          DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS
Sbjct:      400 DQLMAHSHLRYKGTLSMLQCNVFPGLPPDFLDSEVNLFLVPFMDSEAESENPPRAGPGSS 459

Query:      452 PLFSLLLPGYRGHPFSQSLVSKLRQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA 511
          PLFSLLLPGYRGHPFSQSLVSKLRQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA
Sbjct:      460 PLFSLLLPGYRGHPFSQSLVSKLRQVMSMARPQLSHTILTEKNWFHYAARIWDGVRKSSA 519

Query:      512 LAEYSRLLA 520
          LAEYSRLLA
Sbjct:      520 LAEYSRLLA 528

```

Pedant information for DKFZphfbr2 22h13, frame 3

Report for DKFZphfbr2 22h13.3

```
[LENGTH]          520
[MW]               57650.81
[pI]               6.52
[HOMOL]            TREMBL:AC004780_1 product: "F17127_1"; Homo sapiens chromosome 19, cosmid
F17127, complete sequence. 0.0
[PROSITE]          ATP_GTP_A      1
[PROSITE]          MYRISTYL      8
[PROSITE]          CAMP_PHOSPHO_SITE 1
[PROSITE]          CK2_PHOSPHO_SITE 8
[PROSITE]          GLYCOSAMINOGLYCAN 1
[PROSITE]          PKC_PHOSPHO_SITE 3
[PROSITE]          ASN_GLYCOSYLATION 2
[KW]               TRANSMEMBRANE 1
[KW]               LOW COMPLEXITY 11.73 %
```

```
SEQ      MSEGSHSQGLYGIERRRRWKEPGSGGPNLSPGGGRERDYIAPWERRRDASEETSTSV
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccceeeehhhhhhhhcccccee
MEM
```

```
SEQ      MQKPTIILSKPPAERSKQPPPTAPAAPPAPELEKPIVLMPREEGKGPFVAVTGASTPE
SEG      .....XXXXXXXXXXXXXXXXX.....
PRD      eeeeeeeccccccccccccccccccccccccccccceeeccccccccceeeccccc
MEM
```

[illegible]

```
SEQ      IKLVDDQMNMWCDAIEYLLDQTDVLVVGVGLGQTGKSMVMSLLSANTPEEDQRTYVFRA
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      hhhhhccccchhhhhhhhhccccceeeeeccccccchhhhhhhccccchhhhhheeee
MEM
```

```
SEQ      QSAEMKERGGNQTSIGDFFITQERIVFLDTQPILSPSILDHLINNRKLPPEYNLPHTYV
SEG
PRD      hhhhhhhccccceeeeeeeecceeeeeccccccccccccccccccccccccccchh
MEM
```

SEQ EMQSLQIAAFLFTVCHVVI VVQDWFTDLSLYRFLQTAEMVKPSTPSPSHESSSSSSGSDEG
 SEQXXXXXXXXXXXXXXXXXXXXX.....

```

PRD      hhhhhhhhhhhhhhhheeeeeecchhhhhhhhhhhhhcccccccccccccccccc
MEM      MMMMMMMMMMMMMMMMMMMMM.....

SEQ      TEYYPHLVFLQNKARREDFCPRKLRQMHLMIDQLMAHSHLRYKGTLSMLQCNVFPGLPPD
SEG
PRD      cccccceehhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc
MEM      .....

SEQ      FLDSEVNLFLVPMFMDSEAESENPPRAGPGSSPLFSLPGYRGHPSTFQSLVSKLRSQVMSH
SEG
PRD      chhhhhheeeccccccccccccccccccccccccccccccccchhhhhhhhhhhhhh
MEM      .....

SEQ      ARPQLSHTILTEKNWFHYAARIWDGVRKSSALAEYSRLLA
SEG
PRD      hhhhhhhheeeccchhhhhhhhhhhhhcchhhhhhhhhccc
MEM      .....

```

Prosites for DKFZphfbr2_22h13.3

PS00001	30->34	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00002	32->36	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	507->511	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	215->218	PKC_PHOSPHO_SITE	PDOC00005
PS00005	491->494	PKC_PHOSPHO_SITE	PDOC00005
PS00006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	254->258	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	298->302	CK2_PHOSPHO_SITE	PDOC00006
PS00006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS00006	436->440	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	211->217	MYRISTYL	PDOC00008
PS00008	214->220	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008
PS00008	505->511	MYRISTYL	PDOC00008
PS00017	211->219	ATP_GTP_A	PDOC00017

(No Pfam data available for DKFZphfbr2_22h13.3)

DKFZphfbr2_22i4

group: brain derived

DKFZphfbr2_22i4.1 encodes a novel 228 amino acid protein with similarity to the N-terminus of human p52rIPK.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Human P52rIPK N-terminus

complete cDNA, complete cds, few EST hits
function of P52rIPK, repressor of p58IPK protein kinase inhibitor
upstream regulator of interferon induced proteins

Sequenced by AGOWA

Locus: unknown

Insert length: 4748 bp
Poly A stretch at pos. 4726, polyadenylation signal at pos. 4709

```
1 TGGGTCCGGT CCTAGGGTCA CACCCACCGC AGGGTCTGGC TTGGTACAGT
51 TGGGTGCATG CAGAAGTAGG TGGAGCTGCT GTTGCAGCCT TGAGAGAGTT
101 TTATTGTAAA ACTCTTGTA TTTATAGTAA TCGGAGGGGA AAACACCTCT
151 TCCTTTTAAAT TGCTCTGAGG ACCGCTGCCA AAGAAACGCA GTAGATCCGC
201 TCCTCTTTGG GGGCGGGGAG AAAGAACGGG TTGTGTCCGC CATGTTGGTG
251 AAGTCAAGCG AAGGCGACTA GAGCTCCAGG AGGGCCAGTT CTGTGGGCTC
301 TAGTCGGCCA TATTAATAAA GAGAAAGSGA AGGCTGACCG TCCTTCGCCT
351 CCGCCCCCAC ATACACACCC CTTCTTCCCA CTCGCTCTC ACGACTAAGC
401 TCTCAGGATT AAGGCACGCC TGCCTCGATT GTCCAGCCTC TGCCAGAAGA
451 AAGCTTAGCA GCCAGCGCCT CAGTAGAGAC CTAAGGGCGC TGAATGAGTG
501 GGAAGGGGAA ATGCCGACCA ATTGCGCTGC GGGCGGCTGT GCCACTACCT
551 ACAACAAGCA CATTAACTC AGCTTCCACA GGTTCCTTT GGATCCTAAA
601 AGAAGAAAAG AATGGGTTCC CTTGGTTAGG CGCAAAAATT TTGTGCCAGG
651 AAAACACACT TTTCTTTGTT CAAAGCACTT TGAAGCCTCC TGTTTGACC
701 TAACAGSACA AACTCGACGA CTTAAATGG ATGCTGTTC AACCATTTTT
751 GATTTTGTGA CCCATATAAA GTCTATGAAA CTCAGTCAA GGAATCTTTT
801 GAAGAAAAAC AACAGTTGTT CTCAGCTGG ACCATCTAAT TTAATCAAA
851 ACATTAGTAG TCAGCAAGTA CTACTTGAAC ACAGCTATGC CTTTAGGAAT
901 CCTATGGAGG CAAAAAGAG GATCATTAAA CTGAAAAAAG AAATAGCAAG
951 CTTAAGAAGA AAAATGAAA CTTGCCTACA AAAGGAACGC AGAGCAATC
1001 GAAGATGGAT CAAAGCCACG TGTTTGGTAA AGAATTAGA AGCAATAGT
1051 GTATTACCTA AAGGTACATC AGAACACATG TTACCACTG CTTAAGCAG
1101 TCTTCCTTGG GAAGATTTTA AGATCCTTGA ACAAGATCAA CAAGATAAAA
1151 CACTGCTAAG TCTAAATCTA AAACAGACCA AGAGTACCTT CATTAAATTT
1201 TAGCTTGACAG AGAGCTTGAT GCCTATCCTT CATTCTTTTC AGAAGTAAAG
1251 ATAATTATGG CACTTATGCC AAAATTCATT ATTTAATAAA GTTTTACTTG
1301 AAGTAACATT ACTGAATTTG TGAAGACTTG ATTACAAAAG AATAAAAAAC
1351 TCCATATGGA AATTTTATTT GAAAATGAGT GGAAGTGCCT TACATTAGAA
1401 TTACGGACTT AAAAATTTTG CTAATAAATT GTGTGTTTGA AAGGTGTTTT
1451 TTGTTTTTGT CTTTTTAAAC TACTGTTAAA AGAACAGCTT ATGATAAGTA
1501 ATATGTTTAA CTTAGAGAAG AATTTTTTCC TGTACCAAAG TTGGCATATT
1551 GCATTCTAAA TAAGATGCTA AATAAGAGTT AACCACATT CAACATGACC
1601 TTAATACTGC TGGGTTTTGT ATTAATTAAA TTATAATTGG CACTGTGATT
1651 TGAATAATTT ATAGAAAAAA AGGTACAGGG CAAGTTTTTA AATTAAAACT
1701 TTCTATATTT TGTTTACCA GTAAAAGTGA GCTTATCATG GCCTCTCTCA
1751 TAAGAATGAT TTTAAAAATG GTTGTAATAT ATTTTGAATA TATTGAAATG
1801 TGAAGTACCA TTGAGTCATC CAAACTAGGT AAGGCCTCAA GTACTTTAAA
1851 CTAGTAAAT CTAGTAGCTG ATAATATTCA CCTAAGTAAG TGTGTAAAAA
1901 TAATTACAGG TTCAGGACCT AGCTTAGATA AATGTATACT ACTCTTTTTC
1951 TCATAGTAAA AATCTTACAT TTCCAACCTC AAAATTGGTG CTTCCATATT
2001 TGTGTATAAC CAAAACCTCT AAGTTTTTTT GTTTCTTTT TAACTACTTT
2051 CCAATGCACT ACTATACCTC AGAAATAGTG TATCAATATA GTGGGCTTTT
2101 TTTTCTCTCT TCATAAACCC ACAGTAAAA TTAATCACAG GAAACTACTT
2151 ATATCTTAC ACTTTGTATT GATAACTTAA AATGGCATCA GTTTATCTTA
2201 GACATCAGCT TGCTTTTAT CTCTTTTTT AGTGAGTGAA ATAGAGCAAC
2251 TAGCATGCCT GTGTTCCAG CTACTTGGGA GGCTAAGGTG GGAAGATCAA
2301 TTGAACCTAG GAGGTTGAGG CTATAGTGAG CTGTGATTGC ACGACTGCAC
2351 TCCAGCCTCG GCAATGGAGT GAGACTCCTG TCTCTAAAAC AGCAACAACA
2401 AAAATAAAGC AACCATAGTG CATAAGGGAA ATTAATGTT CCCTATAGAA
2451 ATATGTGTAT GTCTGTGATA AATGCTAATT ATTTTATAAA
2501 ATAAAAGTTC AGAACTATT TATCATTTGC CACTTGAACA ATTAAGGGT
2551 TTGCTTTATT TCACTAATGT TTAATAGGAA CCCTTTGCTT CAAACAGGTT
```

```

2601 TGTGAAATC ATGTAAAAAT TTGTTAATAG AGAATCAAGT TATTTAACTC
2651 AACTTATTTA ATTCAAGCTT GTGATACTAA CATACAAAGG TAGCATAAAC
2701 CAAGTCATAA ATTGCTGTAA TCTTCTCTGT AGAGTAATAG CTACTTCATG
2751 ATTTTTTTAA AAATTTCATT TTTTGTCTAT TTAGATTGCG ATTTGCTTGG
2801 CTCTAGTAA CAATTCTTTT ACAGTATTAG CACTCTCTTT ACTAAGGAAT
2851 GCCTCCCAAG GAAATGCAAA GGTAGGAAAA GTCTCTTAGA ATGCCCATGA
2901 GGTATTTAAA ACAGATATTT ATGAAAAATCT TTTTGTGAAT GTTATAAATC
2951 TTGCTAGTTA TTTTATCTTT ATCTTAAGTA TTAGATGTAG TTCTTTGGAA
3001 TTGTCATTAC ATATTTATTT TTTTCTAGTG TGGTTTCAAA TAACTTTTTG
3051 CCAACATATA ATCATCATCA AACATTTCACT GACCATATCT ATTTTATAAC
3101 TCAAAATAAG TTGGACAAAT AATCATTTTA ATAAAACTA TTTTTTCCAA
3151 GTATAACCACT GTGTCATGTGG TTCACCCCTC ACCCCAGATA CAAAAACATT
3201 ATTTGGTAGT CCCAGTTCCC ATCTACAGTA ATACCTTGAA ACCTTAATTA
3251 ATTTAAAAAA TCATAAAAAAT AAAATATTGT AAAATACAAC AAATTTTGGA
3301 CAAGGTTACT TCATCTTCAT TCATTATTAC CTGACAGTAT TAACTACTA
3351 CTCAATAAAT TTAGAGTAAA CTTTCTGTGT TTTTCCCGT GATTTTCATT
3401 GTGCTGTCTT GACAACATGC TCCAACTCT TTGCATCAAA TTGTTTTATT
3451 AACATACATT TGTCTACCTT AAACTAGCT TTATTCACAG AGAAAGACCT
3501 AAAAGGAGTC TATTAATAATG CTGCTTTCAG TTGATAGTT TTTTTTTTAA
3551 TCACTCTGAC CATAAACTAA CTGAAATTAT AATGGATTTT TTTTCTCTC
3601 CCGGTCACAA CACAGATCTT CTGTTTCAAT GTTCTCTGTC TACTGGGCAC
3651 CAACCTCTAC AAAGAACCAG CCAAGGCTA GGTACTTGAT ATAAAAAGGA
3701 ATATTACATT ATTTCTGCC CTCAAGTTGC TCTATCTCTT GAAAGAAACA
3751 AGTAATATTT ATAATACAAT ATGATAAATG CTACAAAAGA AATAGCTGTA
3801 AAGTCCTTTG GTAAATGCTG TTGAATTGGA ATTCAGTAAG AACTATAAAC
3851 TGTAGACCTT TTTATAATCA AATGCTTTTG TCTTGAACA AAACAGATTC
3901 CTCCTTATAT TGACTTAGCA AAGGAGGTAC AAGGACATTG GCATTTGACC
3951 TGAATTATGG TGTTTTATTG AATGAGCTAT AAGACAACAT TTTTACCCTT
4001 TAAAATGAAC ACTGAACAAA TGTGTTAATG GTATCTTTGT TAAAAGGAAA
4051 ACATAGCTAT AAATAAAATA CTACATCGAA ATCCAGCACT GGAGTTCATT
4101 TGAATTTTGA TATTTTGTGT AAAGTAACAA ACCTATTAAC ACAGATTTT
4151 AAAATAACTC AGAATCGTAT AAAGCACTTT GGTACTTATT TGTCTCTTT
4201 TCCCTTACAT TCTGTGTGGT AGGTGGTATT ATCTCTGATT TACACATGAA
4251 GACATCCTTG TTAATGCAAT TTATTTATTC ATTCGGGCAT TTACTGTGTG
4301 CCAACTTGCA AARGGAATAG AAATGTCTGT GATCTAGATA GTTCTAGATT
4351 GAACATAGAT TTTCTGCCAA CAAATCCTCT CTGCTGTTCA CATTATCCTT
4401 TGTTTAACTG ATGAACCAGG TACTAAAT AGGATAAATC ATGTGCTCTA
4451 GAATATGAAA ATAGTAAGGT CTTTGAGGTC ACTTGATCTT CTCTAAGTAG
4501 ACTTTATAAT ATTGTGTTT ATCTCATTTT TCAATATTAG AATACGGGTA
4551 GATTTTAAAT TTGCTATAAT ATAGGAAATG GTTCATCTTT GTACCAAAAT
4601 ATTGCACTCT TCTGATATT AGACAGTTGG AAACCTTCTA AAATTGAGGA
4651 TTTTGTAGTG TATACTAAAT AATTGCATAT TCAAAAAAAT GTATTCTGAG
4701 TATGGTGATA TTAACATTT TCCCCAAAA AAAAAAATA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

98107671:
 Regulation of interferon-induced protein kinase PKR:
 modulation of P58IPK inhibitory function by a novel protein,
 P52rIPK

Peptide information for frame 1

ORF from 511 bp to 1194 bp; peptide length: 228
 Category: similarity to known protein

```

1 MPTNCAAGC ATTYNKHINI SFHRFPLOPK RRKEWVRLVR RKNFVPGKHT
51 FLCSKHFEAS CFDLTGQTRR LKMDAVPTIF DFCTHIKSMK LKSRNLLKKN
101 NSCSPAGPSN LKSNISSQVQ LLEHSYAFRN PMEAKKRIK LEKEIASLRR
151 KMTKCLQKER RATRRWIKAT CLVKNLEANS VLPKGTSEHM LPTALSSSLPL
201 EDFKILEQDQ QDKTLLSLNL KQTKSTFI

```

BLASTP hits

Entry AF007393_1 from database TREMBL:
 product: "P52rIPK"; Homo sapiens P52rIPK mRNA, complete cds.
 Score = 166, P = 2.5e-11, identities = 40/106, positives = 56/106

DKFZphfbr2_22k3

group: brain derived

DKFZphfbr2_22k3 encodes a novel 538 amino acid protein with weak similarity to extensins.

No informative BLAST results; no predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to extensins

complete cDNA, complete cds, few EST hits
CpG Island in 5' UTR complete cDNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2775 bp

Poly A stretch at pos. 2755, polyadenylation signal at pos. 2718

```

1  GGGGCTGCCC GCGCGCTCCA CGGTGCAGAG CTCTAAGCGC GCGGGCTGGC
51 AGGCTGCGGC GCGTCAAGGT CAGCCTGGAG CTGGGTGGCG GCCTGCCTGG
101 GGGCGGGGGA CCCTACTGGA GCGCCGGGCT GGGGCTCCCG AGCGCTCTCG
151 CCATATTGAA TAGCTTCGAC TGGACCGTCT TTGTCTGCGA AGTCTGTGCC
201 CAAGTTCCAG CCGCGTCCCT GGGGCTGGG GCAGGAAGAG TCGCTGGCAG
251 CCGCGCGGCC CCAACTTGGA GCTGGGACAC CACGTTTCCA GCTTGGAGTG
301 GGCCTTGAGC CTTGGGACTG ACCTCGCCCC CGGCTCACGT AGGCATCCTG
351 GAAATTGATT CCCCAGATC CTGGTGGGG GAGCCGGAAT TGGTCAAGAC
401 TGTACTTGTT GCAGGCGAAG AGATTGGAGG CGTTTGGCTC GTCCCTGGCT
451 AGGGAGGTGA GACTCTCCGG TCAGCGTTGC TGGAACTCCC CCCATCCAGT
501 CCCTCCCTCA AGACTAAGGS CTACAGTAGT TTGTTGGGGC TCATTGCCCC
551 CTCACCCAG ATATCACCTT GGAGATCTTA AAGACTCTCG AGAAAAGCCA
601 CGTGGGGGGC TGGTTCCTCT GGGGCTTCTT GCCCTCCCCG GACTGCCTCA
651 TTCTTTGGAG CGTCCCGCAT GTCTGCAAAAG ATGTGGATTT GGACGTCTCT
701 GTGGAAGCCC TAAAGCCCGT GGGGACATTT AAGAAGATCG GCAAGGTGTT
751 CCGCAAGGAG CAGGACTCCA CGGTGGGGAT GCTGCAGATC GGGGAGGACG
801 TCGACTATTT GCTCATCCCC CGGGAGGTCA GGCTGGCTGG GGGCGTCTGG
851 AGAGTCACTT CTAAGCCCGC CACCAAGGAA GCAGAAATTC GGGAGCGGCT
901 GACCCAGTTC CTGGAAGAAG AGGGCCGAC CCTGGAGGAC GTGGCCCGCA
951 TCATGGAGAA GAGCACCCTG CACCCGCCCC AGCCCCCAA AAAGCCCAAG
1001 GAGCCCCGAG TGAGGAGGAG AGTGCAGCAG ATGGTGACTC CTCGCCCCCG
1051 GCTGTCTGTG GGCACGTACG ACAGCAGCAA CGCCAGCGAC AGCGAGTTCA
1101 GCGACTTCGA GACCTCCAGA GACAAGAGCC GCCAGGGCCC GCGGCGGGGC
1151 AAGAAGGTGC GCAAAATGCC CGTCAGTAC CTGGGCAGCA AGTTCCTGGG
1201 AAGCGACCTG GAGAGTGAGG ATGATGAGGA ACTGGTCGAG GCCTTCCTCC
1251 GCGCAGAGGA GAAGCAGCCC AGCGCGCCCG CTGCCCCGCG CCGCGTCAAC
1301 CTGCCAGTGC CCATGTTTGA GGACAACTG GGGCCTCAGC TGTCCAAAGC
1351 GGACAGGTGG CCGGAGTATG TCAGCCAGGT GTCTGGGGG AAGCTGAAAG
1401 GGAGGGTGAA GGGTTGGGCG CCGAGGGCGG GCCCGGGGTT GGGCGAGGCC
1451 CGGCTGGCCT CCACCGCAGT GGAGAGCGCA GGGGTATCAT CGCGCCAGA
1501 GGGCACAGC CCGGGGATC GCTTGGGAAA CGCGGGAGAT GTTTGTGTGC
1551 CCCAGGCTTC CCCTAGGCGA TGGAGGCCCC AGATCAACTG GGCCTCCTTT
1601 CGGCGCCGCA GGAAGGAGCA GACAGCACCC ACAGGTCAAG GGGCAGACAT
1651 CGAGGCTGAT CAGGGGGGAG AGGCTGCAGA TAGTCAAAGG GAAGAGGCCA
1701 TAGCTGACCA GCGGGAAGGG GCTGCAGGTA ATCAGAGGGC TGGGGCCCCA
1751 GCTGACCAGG GGGCAGAGGC TGCAGATAAT CAGAGGGAAG AGGCTGCAGA
1801 TAATCAGAGG GCAGGGGCCC CAGCTGAGGA GGGGCGAGAG GCTGCAGATA
1851 ACCAGAGGGA AGAGGCTGCA GATAATCAGA GGGCAGAGGC CCCAGCTGAC
1901 CAGAGGTAC AGGGCACAGA TAAACACAGG GAAGAGGCTG CAGATAATCA
1951 GAGGGCGGAG GCCCAGCTG ACCAGGGGTC AGAGGTTACA GATAATCAAA
2001 GGAAGAGGCG CGTACATGAC CAGAGGGAAG GGGCCCCAGC TGTCCAGGGT
2051 GCAGATAATC AGAGGGCACA GGGCCGGGCT GGCCAGAGGG CAGAGGCTGC
2101 ACATAATCAG AGGGCAGGGG CCCCAGGTAT CCAGGAAGCT GAAGTCTCAG
2151 CTGCCCAAGG GACCACAGGA ACAGCTCCAG GAGCCAGGGC CCGGAAACAG
2201 GTCAAGACAG TGAGGTTCCA GACCCCTGGA CGCTTTTCGT GGTTTTGCAA
2251 GCGCCGGAGA GCCTTCTGGC AACTTCCCG GTTGCCAAAC CTGCCCAAGA
2301 GAGTCCCCAG GGCAGGAGAG GTCAGGAACC TCAGGGTGCT GAGGGCCGAG
2351 GCCAGAGCAG AAGCTGAGCA GGGAGAGCAA GAAGACCAGC TGTGAGGTGA
2401 GGGCTAGAGA CAGCCACGG GCCCTCCCTC CAAGTGTGGG AGGGAGAGAT
2451 GCTCTGCCTC TGAACCTCAA AGTGGAGGTG GAGTGTCTGC CACGTCTCCA
2501 CCTAACACCC CTCTTTATTC TCTTGTAAA GTTTTGTGTA TGCTTTGATT
2551 TTTTTTAAA TTTTTTAGAG ACAGGGTCTC ACTCTGTGTC CCAGGCTGGA
2601 GTGCAGTGGC ATGATCATAA CTCAGTGCAG CCTCAAACCT CTGGCCTCAA
2651 GTGATCTCTC TGCTCGGCC TCCCAAAATG CTGGGATTAC AGATGTGAGC

```

2701 CACCACACAC ACCATCTGAT TAAAAAATAA AAATACTGAT TCCCTGTAGC
 2751 AACCCAAAAA AAAAAAATAA AAAAA

BLAST Results

Entry HS164A7F from database EMBL:
 H.sapiens CpG island DNA genomic MseI fragment, clone 164a7, forward
 read cpgl64a7.ftla .
 Score = 740, P = 3.0e-25, identities = 150/151

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 779 bp to 2392 bp; peptide length: 538
 Category: similarity to known protein

1 MLQIGEDVDY LLIPREVRLA GGVWRVISKP ATKEAEFRER LTQFLEEEGR
 51 TLEDVARIME KSTPHPPQPP KKPKEPRVRR RVQQMVTPPP RLVVGTYDSS
 101 NASDSEFSD FTSRDKSRQG PRGKKVRKM PVSYLGSKFL GSDLESEDDE
 151 ELVEAFLRRQ EKQPSAPPAR RRVNLPVPMF EDNLGPQLSK ADRWREYVSQ
 201 VSWGKLRKRV KGWAPRAGPG VGEARLASTA VESAGVSSAP EGTSPGDRLG
 251 NAGDVCVPQA SPRRWPKIN WASFRRRRKE QTAPTGGQAD IEADQGGEAA
 301 DSQREEAIAD QREGAAGNQR AGAPADQGAE AADNQREEAA DNQRAGAPAE
 351 EGAEAAADNQR EEAADNQRAE APADQRSQGT DNHREEAADN QRAEAPADQC
 401 SEVTDNQREE AVHDQREERAP AVQGADNQRA QARAGQRAEA AHNQRAGAPG
 451 IQEAEVSAAQ GTTGTAPGAR ARKQVKTVERF QTPGRFSWFC KRRRAFHWTP
 501 RLPTLPKRVP RAGEVRNLRV LRAEARAEEAE QGEQEDQL

BLASTP hits

Entry RNU67136_1 from database TREMBL:
 "A-kinase anchoring protein AKAP150"; Rattus norvegicus
 A-kinase anchoring protein AKAP150 mRNA, complete cds. Rattus
 norvegicus (Norway rat)
 Length = 714
 Score = 182 (64.1 bits), Expect = 1.2e-10, P = 1.2e-10
 Identities = 73/257 (28%), Positives = 104/257 (40%)

Alert BLASTP hits for DKFZphfbr2_22k3, frame 2

TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds., N = 1, Score = 178, P = 3.7e-11

>TREMBL:PFSANTY_1 product: "S-antigen"; Plasmodium falciparum KF1916
 S-antigen gene, complete cds.
 Length = 285

HSPs:

Score = 178 (26.7 bits), Expect = 3.7e-11, P = 3.7e-11
 Identities = 60/217 (27%), Positives = 97/217 (44%)

Query: 269 INWASFRRRRKEQTAPTGGQA-DIEADQGGEAADSQRE-EAIADQ---REGAAGNQAGA 323
 +N + + + E G+G D E E +D+ E E I Q E A N+ AG+
 Sbjct: 47 LNGKNGKGNKYEDLQEEGEGENDDEHSNSEESDNDEENEIIVGQDGSNEKAGSNEEAGS 106

Query: 324 PADQGAEEADNQREEAADNQAGAPAEEGA--EAADNQR----EEAADNQRAEAPADQRS 377
 G+ E+A N++AG+ E G+ EA N+ EEA N++A + S
 Sbjct: 107 NEKAGSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGS 166

Query: 378 QGTDNHREEAADNQRAEAPADQGEVTDNQREEAVHDQREERAPAVQGADNQRAQAR--AG 435
 EEA N++A + + GS E+A +++ + G+ N++A + AG
 Sbjct: 167 NEKAGSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEEAGS-NEKAGSNEEAG 225

Query: 436 QRAEAAHNQRAQA---PGIQEAEVSAAQGTGTGA-PGA 469

EA N+ AG+ G E + +G GT PG+

Sbjct: 226 SNEEAGSNEEAGSNEEAGSNEGSEAGTEGPKGTGGPGS 263

Score = 173 (26.0 bits), Expect = 1.5e-10, P = 1.5e-10
Identities = 51/190 (26%), Positives = 83/190 (43%)

Query: 279 KEQTAPTQ-GADIEADQGGEAADSQREEAIADQREGAAGNQAGAPADQGAEEADNQRE 337
+E QQ G++ +A EA +++ A E A N++AG+ G+ E

Sbjct: 83 EENEIIVGQDGSNEKAGSNEEAGSNEK-----AGSNEEAGSNEKAGSNEEAGSNE 138

Query: 338 EAADNQAGAPAEAGAEADNQREAAADNQRAEAPADQRSQGTDNHREAAADNQRAEAPA 397
EA N+ AG+ E G+ E+A N++A + + S EEA N++A +

Sbjct: 139 EAGSNEEAGSNEEAGSNEKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEEAGSNEKAGSNE 198

Query: 398 DQGSSEVTDNQREEAVHDQREAPAVQAGADNQRAQARAGRAEAAHNQAGAPGIEAEVS 457
GS EEA +++ + G++ + AG EA N+ AG+ EA

Sbjct: 199 KAGSNEKAGSNEEAGSNEKAGSNEEAGSNEE-----AGSNEEAGSNEEAGSNEGSEAGTE 253

Query: 458 AAQGTGTGAPG 468
+GT G G

Sbjct: 254 GPKGTGGPGSG 264

Score = 147 (22.1 bits), Expect = 1.6e-07, P = 1.6e-07
Identities = 40/168 (23%), Positives = 70/168 (41%)

Query: 288 GADIEADQGGEAADSQR--EEAIADQREGAAGNQAGAPADQGAEEADNQREAAADNQRA 345
G++ EA +A +++ A E A N+ AG+ + G+ E+A N++A

Sbjct: 111 GSNEEAGSNEKAGSNEKAGSNEEAGSNEEAGSNEEAGSNEEAGSNEKAGSNEKAGSNEKA 170

Query: 346 GAPAEAGAEADNQREAAADNQRAEAPADQRSQGTDNHREAAADNQRAEAPADQGSSEVTD 405
G+ E G+ EEA N++A + S EEA N++A + + GS

Sbjct: 171 GSNEEAGSNEKAGSNEEAGSNEKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEEAGSNEEA 230

Query: 406 NQREEAVHDQR--ERAPAVQAGADNQRAQARAGRAEAAHNQAGAPGI 451
EEA ++ + G + + G E +HN++ I

Sbjct: 231 GSNEEAGSNEEAGSNEGSEAGTEGPKGTGGPGSGGESHKKKSKKSI 278

Score = 101 (15.2 bits), Expect = 2.5e-02, P = 2.4e-02
Identities = 26/100 (26%), Positives = 47/100 (47%)

Query: 281 QTAPTGGADIEADQGGEAADSQREEAIADQREGAAGNQAGAPADQGAEEADNQREAAA 340
+ A + + A + G EEA +++ G+ N++AG+ G+ E+A

Sbjct: 162 EKAGSNEKAGSNEEAGSNEKAGSNEEAGSNEKAGS--NEKAGSNEKAGSNEEAGSNEKAG 219

Query: 341 DNQAGAPAEAGAEADNQREAAADNQRAEAPADQRSQGT 380
N+ AG+ E G+ EEA N+ +EA + +GT

Sbjct: 220 SNEEAGSNEEAGSNEEAGSNEEAGSNEGSEA-GTEGPKGT 258

Pedant information for DKF2phfbr2_22k3, frame 2

Report for DKF2phfbr2_22k3.2

[LENGTH] 538
[MW] 59402.19
[pI] 8.72
[HOMOL] TREMBL:AF037364_1 gene: "MA1"; product: "paraneoplastic neuronal antigen MA1";
Homo sapiens paraneoplastic neuronal antigen MA1 (MA1) mRNA, complete cds. 4e-10
[PROSITE] AMIDATION 1
[PROSITE] MYRISTYL 12
[PROSITE] CK2_PHOSPHO_SITE 11
[PROSITE] PKC_PHOSPHO_SITE 6
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 18.03 %

SEQ MLQIGEDVDYLLIPREVLAGGVWRVISKPATKEAEFRERLTQFLEEEGRTLEDVARIME
SEG
PRD ccc

SEQ KSTPHPPQPPKKPKPRVRRRVQQMVTPPPRLVVGTYDSSNASDSEFSDFETS RDKSRQG
SEGxxxxxxxxxxxxxxxxxxxxx.....
PRD hcc

SEQ PRRGKKVRKMPVSYLGSKFLGSDLESEDEELVEAFLRRQEKQPSAPPARRRVNLPVPMF
SEGxxxxxxxxxxxxxxxxxxxxx.....
PRD ccc

```

SEQ      EDNLGPQLSKADRWREYVSQVSWGKLRVRKGWAPRAGPGVGGEARLASTAVESAGVSSAP
SEG      .....
PRD      cccccccchhhhhhhheeeccchhhhhccccccccccchhhhhhhhhcccccc

SEQ      EGTSPGDRNLGNAGDVCVPQASPRRWRPKINWASFRRRRKEQTAPTGGGADIEADQGGEEA
SEG      .....
PRD      cccccccccccccceeeccccccccccchhhhhhhhhhhccccccchhhhhccchhh

SEQ      DSQREEAIADQREGAAGNQAGAPADQGAEEADNQREEADNQAGAPAEEGAEADNQ
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhh

SEQ      EEAADNQRAEAPADQRSQGTDNHREEADNQRAEAPADQGEVTDNQREEAVHDQREAP
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      AVQGADNQRAQARAGQRAEAAHNQAGAPGIQAEVSAAQGTGTAPGARARKQVKTVRF
SEG      .....
PRD      hccccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhhhhh

SEQ      QTPGRFSWFCKRRRAFHWHTPRLPTLPKRVPRAGEVRNLRLVRAEAREAEQGEQEDQL
SEG      .....
PRD      cccccceehhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcc

```

Prosites for DKFZphfbr2_22k3.2

PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS00005	273->276	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	499->502	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	112->116	CK2_PHOSPHO_SITE	PDOC00006
PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	189->193	CK2_PHOSPHO_SITE	PDOC00006
PS00006	229->233	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	220->226	MYRISTYL	PDOC00008
PS00008	242->248	MYRISTYL	PDOC00008
PS00008	296->302	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	328->334	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	450->456	MYRISTYL	PDOC00008
PS00008	461->467	MYRISTYL	PDOC00008
PS00008	464->470	MYRISTYL	PDOC00008
PS00009	123->127	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_22k3.2)

DKFZphfbr2_22k8

group: brain derived

DKFZphfbr2_22k8 encodes a novel 172 amino acid protein without similarity to known proteins.

No informative BLAST results; no predictive prosite, pfam or SCOP motif

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: /map="7"

Insert length: 2789 bp

Poly A stretch at pos. 2769, polyadenylation signal at pos. 2756

```
1 GGGGGAGCCA TGAGGCGCCA GCCTGCGAAG GTGGCGGCGC TGCTGCTCGG
51 GCTGCTCTTG GAGTGCACAG AAGCCAAAAA GCATTGCTGG TATTTGCAAG
101 GACTCTATCC AACCTATTAT ATATGCCGCT CCTACGAGGA CTGCTGTGGC
151 TCCAGGTGCT GTGTGCGGGC CCTCTCCATA CAGAGGCTGT GGTACTTCTG
201 GTTCCTTCTG ATGATGGGCG TGCTTTTCTG CTGCGGAGCC GGCTTCTTCA
251 TCCGGAGGCG CATGTACCCC CCGCCGCTGA TCGAGGAGCC AGCCTTCAAT
301 GTGTCTTACA CCAGGCAGCC CCCAAATCCC GGCCGAGGAG CCCAGCAGCC
351 GGGGCCGCCC TATTACACTG ACCCAGGAGG ACCGGGGATG AACCCGTGCG
401 GGAATTCACG GGCAATGGCT TTCCAGGTCC CACCCAACCT ACCCCAGGGG
451 AGTGTGGCCT GCCCGCCCCC TCCAGCCTAC TGCAACACGC CTCGCCCCCC
501 GTACGAACAG GTAGTGAAGG CCAAGTAGTG GGGTGCCACG GTGCAAGAGG
551 AGAGACAGGA GAGGGCCTTT CCTTGGCCTT TCTGTCTTCG TTGATGTTCA
601 CTTCCAGGAA CGGTCTCGTG GGCTGCTAAG GGCAGTTCCT CTGATATCCT
651 CACAGCAAGC ACAGCTCTCT TTCAGGCTTT CCATGGAGTA CAATATATGA
701 ACTCAGACTT TGTCTCCTCT GTTGCTTCTG TTTCTGACGC AGTCTGTGCT
751 CTCACATGGT AGTGTGGTGA CAGTCCCGGA GGGCTGACGT CCTTACGGTG
801 GCGTGACACG ATCTACAGGA GAGAGACTGA GAGGAAGAAG GCAGTGCTGG
851 AGGTGCAGGT GGCATGTAGA GGGGCCAGGC CGAGCATCCC AGGCAAGCAT
901 CCTTCTGCCC GGGTATTAAAT AGGAAGCCCC ATGCCGGGCG GCTCAGCCGA
951 TGAAGCAGCA GCCGACTGAG CTGAGCCGAG CAGGTCATCT GCTCCAGCCT
1001 GTCCTCTCGT CAGCCTTCCT CTTCCAGAAG CTGTTGGAGA GACATTACAG
1051 AGAGAGCAAG CCCCTTGTCG TGTTTCTGTC TCTGTTTATA TCCTAAAGAT
1101 AGACTTCTCC TGCAACGCCA GGGGAAGGATA GCACGTGACG CTCTCACCAG
1151 AGGATGGGGC CTAGAATCAG GCTTGCCTTG GAGGCGTGAC AGTGATCTGA
1201 CATCCACTAA GCAAAATTTAT TTAATTCAT GGGAAATCAC TTCCTGCCCC
1251 AACTGAGAC ATTGCATTTT GTGAGCTCTT GGTCTGATTT GGAGAAAGGA
1301 CTGTTACCCA TTTTTTTGGT GTGTTTATGG AAGTGCATGT AGAGCGTCTT
1351 GCCCTTTGAA ATCAGACTGG GTGTGTGTCT TCCCTGGACA TCACTGCCCT
1401 TCCAGGGCAT TCTCAGGCCC GGGGGTCTCC TTCCTCAGG CAGCTCCAGT
1451 GGTGGGTCTT GAAGGGTGCT TTCAAACCG GGCACATCTG GCCCGGAAGT
1501 CACATGGACT CTTCCAGGGA GAGAGACCA GAGGCGGCTC TCTCTCTGAG
1551 GTTGTGTTGG GTCTAAGCGG GTGTGTGCTG GGCTCCAAGG AGGAGGAGCT
1601 TGCTGGGAAA AGACAGGAGA AGTACTGACT CAACTGCACT GACCATGTTG
1651 TCATAATTAG AATAAAGAAG AAGTGGTCTG AATGCACAT TCCTGGATAG
1701 GAATCACAGC TCACCCAGG ATCTCACAGG TAGTCTCCTG AGTAGTTGAC
1751 GGCTAGCGGG GAGCTAGTTC CGCCGCATAG TTATAGTGTT GATGTGTGAA
1801 CGCTGACCTG TCCTGTGTGC TAAAGAGTAT GCAGCTTAGC TGAGGCGCCT
1851 AGATTACTAG ATGTGCTGTA TCACGGGGAA TGAGGTGGGG GTGCTTATTT
1901 TTTAATGAAC TAATCAGAGC CTCTTGAGAA ATTGTACTC ATTGAATGG
1951 AGCATCAAGA CATCTCATGG AAGTGGATAC GGAGTGATTT GGTGCCATG
2001 CTTTCACTC TGAGGACATT TAATCGGAGA ACCCTCTGGG GAATTTTGTG
2051 GGAGACACTT GGGAAACAAA CAGACACCTT GGGAAATGCAG TTGCAAGCAC
2101 AGATGCTGCC ACCAGTGTCT CTGACCACCC TGGTGTGACT GCTGACTGCC
2151 AGCGTGGTAC CTCCCATGCT GCAGGCTCCT ATCTAAATGA GACAAACAA
2201 CACAATGTTT ACTGTTTACA ACCAAGACAA CTGCGTGGGT CCAAACACTC
2251 CTCCTCCCTC AGGTCATTTG TTTTGCATTT TTAATGTCTT TATTTTGTG
2301 AATGAAAAAG CACACTAAGC TGCCCCGGA ATCGGGTGCA GCTGAATAGG
2351 CACCCAAAAG TCCGTGACTA AATTCCGTTT GTCPTTTTGA TAGCAAAATA
2401 TGTTAAGAGA CAGTGATGCG TAGGGCTCAA CAATTTTGTA TTCCCATGTT
2451 TGTGTGAGAC AGAGTTTGTG TCCCTTGAA CTTGGTTAGA ATTGTGCTAC
2501 TGTGAACGCT GATCCTGCAT ATGGAAGTCC CACTTTGGTG ACATTTCTCG
2551 GCCATTCTTG TTTCCATTGT GTGGATGGTG GGTGTGCCCC ACTTCTGGA
2601 GTGAGACAGC TCCTGCTGTG TAGAATCCCC GAGAGCTCCG TGGTTCAGAG
2651 TAAACTTGAA GCAGATCTGT GCATGCTTTT CCTCTGCAGC AATTGGCTCG
2701 TTTCTCTTTT TTGTTCTCTT TTGATAGGAT CCTGTTTCCT ATGTGTGCAA
```

BLAST Results

Medline entries

Peptide information for frame 1

BLASTP hits

```

Query:      147 GSVACPPPPPAYCNTPPPP 164
           S   PPPP Y N P PP
Sbjct:     98 YSSPPPPPPFYENIPLPP 115

  Score = 95 (14.3 bits), Expect = 2.4e-04, P = 2.4e-04
  Identities = 24/61 (39%), Positives = 29/61 (47%)

Query:     104 PPNPGPGAQQPGPPYYTDPGGGPMNPNVGNSTAMAFQVPPNSPQGSVACPPPPPAYCNTPPPP 163
           PP+P P   P P YY P P +P      ++ PP P S   PPPP Y +PPP
Sbjct:     1  PPSFSP----PPPYYYKSPPPPSPSP---PPPYYYKSPP-PPSPS---PPPPYYKSPPP 49

Query:     164 P 164
           P
Sbjct:     50 P 50

  Score = 68 (10.2 bits), Expect = 4.2e+00, P = 9.8e-01
  Identities = 24/69 (34%), Positives = 29/69 (42%)

Query:     87 PPLIEEPAFNVSYSYTRQPP---NPGPGAQQPGPPYYTDPGGGPMNPNVGNSTAMAFQVPPN 13
           PPP   P   Y PP   +P + P PP Y+ P P   P   + +   PP
Sbjct:     63 PPPDPSPPPPPYYKSPPPPSPSPPPSPSPPPPTSYSSPPPPP--PFYENIPL----PPV 116

Query:     144 SPQGSVACPPPP 155
           S A PPPP
Sbjct:     117 IGV-SYASPPPP 127

```

Peptide information for frame 3

ORF from 0 bp to 368 bp; peptide length: 123
Category: questionable ORF
Classification: unset

1 GSHEAPACEG GGAAARAALG VHRSQALLV FRRTLSNLLY MPLLRGLLWL
51 QVLCAGPLHT EAVVLLVPSD DGRAFLRLSR LLHPEAHVPP AADRGASLQC
101 VLHOAAPKSR PRSPAAGAAL LH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 22k8, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 22k8, frame 1

Report for DKF2phfbr2_22k8.1

[illegible]

(No Prosite data available for DKFZphfbr2 22k8.1)

(No Pfam data available for DKFZphfbr2_22k8.1)

Pedant information for DKFZphfbr2_22k8, frame 3

Report for DKFZphfbr2_22k8.3

[LENGTH] 122
[MW] 12854.08
[pI] 10.27
[KW] All_Alpha
[KW] LOW_COMPLEXITY 25.41 %

SEQ GSHEAPACEGGGAAARAALGVHRSQKALLVFRRTLSNLLYMPLLRGLLWLQVLCAGPLHT
SEGXX
PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhccccccccchhhhhhhcccccc

SEQ EAVVLLVPSDDGRAFLRLSRLLHPEAHVPPAADRGLQCVLHQAPKSRPRSPAAGAAL
SEGXX
PRD cceeeeeccccchhhhhhhccccccccccccccccchhhhhhhccccccccchhhhhc

SEQ LH
SEG ..
PRD cc

(No Prosite data available for DKFZphfbr2_22k8.3)

(No Pfam data available for DKFZphfbr2_22k8.3)

DKFZphfbr2_23b10

group: nucleic acid managment

DKFZphfbr2_2b10 encodes a novel 580 amino acid protein with strong similarity to rat RNA helicase HEL117.

HEL117 is a DEAD/H box helicase, which co-localises with a splicing factor and thus seems to be involved in splicing.

The new protein can find application in modulation of splicing.

strong similarity to rat RNA helicase HEL117

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2905 bp

Poly A stretch at pos. 2885, no polyadenylation signal found

```

1  GGGGGCTCCG CTCCGCACCA CCAACCCCGG GCCGCAGTCC TGACGAGCGG
51  GTCAGGGCTT GTCGGGCGGA AGCCTGGCCT GGAGCCTGGA AGGGGGAGAC
101 GGCCCGAGCG GGAGCGGGAG CGGACCGCGC CTCAGTCCTG CGCGGAATAT
151 TGAAGGATGT TTGTTCCAAG ATCTCTAAAA ATCAAGAGGA ATGCTAATGA
201 TGATGGCAAA AGTTGTGTGG CTAAGATAAT TAAACCAGAC CCAGAAGACC
251 TTCAGTTGGA CAAAAGCAGA GATGTTCCCG TTGATGCTGT AGCTACAGAA
301 GCAGCCACAA TAGACAGGCA CATCAGCGAA TCATGCCCTT TCCCCAGCCC
351 AGGTGGCCAG TTGGCAGAGG TTCATTGAGT AAGTCCCGAG CAGGGTGCGA
401 AGGACAGCCA TCCTTCTGAA GAGCCCGTTA AGTCATTTTC CAAAACACAG
451 CGCTGGGCAG AACCAGGGGA ACCCATCTGT GTTGTCTGTG GTCGTTATGG
501 AGAGTATATC TGTGATAAGA CAGATGAAGA TGTGTGTAGT TTGGAGTGA
551 AAGCGAAACA TCTTCTACAA GTTAAGGAAA AGGAAGAGAA ATCAAAACTC
601 AGCAATCCAC AGAAGGCTGA TTCTGAGCCA GAGTCTCCAC TGAATGCTTC
651 CTATGCTTAC AAAGAGCACC CCTTTATTTT GAACCTTCAG GAAGACCAGA
701 TTGAAATCTT TAAACAGCAG CTGGGAATTT TAGTTCGAAG GCAAGAAGTC
751 ACCAGGCCCA TTATTGACTT TGAACATTGT AGTCTCCCTG AGGTCTTAAA
801 TCACAATCTG AAGAAATCAG GCTATGAGGT GCCAACTCCC ATTCAAATGC
851 AGATGATTCC TGTGGGACTT CTGGGAAGAG ACATTCTGSC CAGTGCAGAT
901 ACTGGCTCAG GAAAAACAGC TGCTTTTCTT CTTCCTGTTA TCATGCGAGC
951 TTTATTCGAG AGCAAACTC CATCTGCGCT CATTCCTTACA CCAACCAGAG
1001 AGTTAGCCAT TCAGATAGAG AGACAAGCTA AAGAATTGAT GAGTGGCCTG
1051 CCACGCATGA AAACGTGTCT TCTTGTAGGG GGCTTACCCT TACCCCCACA
1101 GCTTTATCGT CTGCAACAAC ATGTTAAGGT TATCATAGCA ACCCTGGGCG
1151 GACTTCTGGA TATAATAAAG CAGAGCTCTG TAGAACTCTG TGGTGTAAAG
1201 ATTGTGGTAG TAGATGAAGC TGATACCATG TTAAGATGCG GTTTTCAACA
1251 ACAAGTGCTT GACATTTTGG AAAACATTCC TAATGATTGT CAGACCATTT
1301 TGGTTTCAGC CACAAATCCA ACTAGCATAG AACAGTAGC AAGCCAGCTT
1351 CTGCATAATC CTGTGAGAAT TATCACTGGA GAAAAGAAC TACCTTGTGC
1401 CAATGTACGT CAGATTATTT TGTGGGTAGA AGACCCAGCC AAAAAGAAAA
1451 AATTATTTGA AATTTTAAAT GATAAGAAAC TCTTTAAGCC TCCAGTGTTA
1501 GTATTTGTGG ACTGCAAACT AGGAGCAGAT CTTTGTAGTG AAGCCGTTCA
1551 GAAATACACA GGGCTGAAAA GCATATCTAT ACATTGCGAG AAGTCGCAAA
1601 TAGAAAGGAA AAACATATTG AAGGGATTAC TTGAAGGAGA CTATGAAGTT
1651 GTAGTGAGCA CAGGAGTCTT GGGACGAGGC CTAGACTTGA TCAGTGTCAG
1701 GCTGGTTGTC AATTTTGATA TGCCTTCAAG TATGGATGAG TATGTCCATC
1751 AGGAAATATC CTACAAGTCT ACTTGGAGGA ATCCCAGCA TTTTCAACAG
1801 GATGTCAGAA TGACCTTGGG CTATGTTGGC AAAGCACAA GGGAAAGAAG
1851 CAACCAATTG AAGTCAAAAC TAGGCCTTAA AAAAAATTGT TCTTCTAAA
1901 TGAAACTTTA TGAAGACCC AAGCTTCCTT TATGTAAAAA TAGGATACTC
1951 ACTAGGCTTT GGGGCTGACA ATGGTTTTTA AATCTTGCTA ATCTTCCCTG
2001 GAATGAAACC AGCATGACTC AAAGAGAAAA AGAGAGTCTA TAATATTTTC
2051 TAATCCCTGA GTTCTTTTCT TTATATATTA AAAAGGATTA TTAGGCTGGG
2101 TGTGGTGGCT CACGCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGGGAGT
2151 GGATCACCTG AGTTCGAGAC CAGCCTAACC AACATGGAGA AACCCTGTCT
2201 CTAATAAAAA TACAAAATTA GCCAGGCGTG GTGGCGCATG CCTGTAATCC
2251 CAGCTACTCA GGAGGCTACA GCAGGAGAA TGTGTAAGT CCGGAGGCAG
2301 AGCCAAGATC GCACCACTGC ACTCCAGCCT GGGCAACAAG AAGTGAACCT
2351 TGTCTCAAAA TAATATTAAT GATAATAATA ATAATAATAA TAGGGATTAC
2401 TTGCATAATT GTTCTTTTAA AATTATTGGC AGTATTGCTG AATGTATTTA
2451 GATTTTTCAC CCAAGTGACA ACAACTGAAT TCATAAAGAT TCATCAACAA
2501 GACCTGATAA AAAAAATGT AAGCATATTA TAGTGGATAC TTCCAAGACT
2551 CTTGGTCTAA CATGTATTAG AAAGCAGAAG GAGCCAGGC ACAGGGGCTC
2601 CCGCCGGTAA TCCCAAAGCT TTGGGAAGCC AAGGCAGGTG GATCGCTTGA
2651 GCTCAGGAGT TAGAGACCAG CCTGGGCAAC ATGGTGAAT CCCGTACCA

```

2701 CAAAAAATG CAAAAATTAA CTGGGCGTGG TGGCATGCAC CTGTAGTCCC
 2751 AGCTACTCTG GAGGCTGAGG TGAGGGGAAT CACCTGAGCC GGGGGAATCA
 2801 CCTGAGCCCA GGAAGTTGA GGCTGCTGTG AGCCATGGTC ATGACACTGC
 2851 CCTCCAGCCT GGACAACAGA TTGAGACCCT GTCTCAAAA AAAAAAAAAA
 2901 AAAAA

BLAST Results

No BLAST result

Medline entries

Medline:

A putative mammalian RNA helicase with an arginine-serine-rich domain

Peptide information for frame 1

ORF from 157 bp to 1896 bp; peptide length: 580
 Category: strong similarity to known protein
 Prosite motifs: ATP_GTP_A (247-255)
 LEUCINE_ZIPPER (298-320)

1 MFVPRSLKIK RNANDDGKSC VAKIIPDPE DLQDKSRDV PVDVATEAA
 51 TIDRHISESC PFPSPGGQLA EVHSVSPEQG AKDSHPSEEP VKSFSKTQRW
 101 AEPGEPICVV CGRYGEYICD KTDDEVCSE CKAKHLLQVK EKEEKSLSN
 151 PQKADSEPE PLNASYVYKE HPFILNLQED QIENLKQQLG ILVQGQEVTR
 201 PIIDFEHCSL PEVLNHNKKG SGYEVPTPIQ MQMIPVGLLG RDILASADTG
 251 SGKTAFFLLP VIMRALFESK TPSALILTPT RELAIQIERQ AKELMSGSLPR
 301 MKTVLLVGGGL PLPPQLYRLQ QHVKVIIATP GRLLDIIKQS SVELCGVKIV
 351 VVDEADTMLK MGFGQQVLDI LENIPNDCQT ILVSATIPTS IEQLASQLLH
 401 NPVRIITGEK NLPCANVRQI ILWVEDPAKK KKLFEILNDK KLFKPPVLVF
 451 VDCKLGADLL SEAVQKITGL KSISIHSEKS QIERKNILKG LLEGDYEVVV
 501 STGVLGRLGD LISVRLVVNF DMPSSMDEYV HQENTYKSTW RNPQHQQQDV
 551 RMTLGYVGKA QWEEDNQLKV KLGLKKNCS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23b10, frame 1

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 615, P = 1.6e-60

TREMBL:AB018344_1 gene: "KIAA0801"; product: "KIAA0801 protein"; Homo sapiens mRNA for KIAA0801 protein, complete cds., N = 1, Score = 615, P = 2.8e-59

TREMBL:CEFO1F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1., N = 2, Score = 365, P = 1.9e-58

TREMBL:AF083255_1 product: "RNA helicase-related protein"; Homo sapiens RNA helicase-related protein mRNA, complete cds., N = 2, Score = 556, P = 1.5e-57

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces pombe), N = 1, Score = 591, P = 1.6e-57

>PIR:A57514 RNA helicase HEL117 - rat
 Length = 1,032

HSPs:

Score = 615 (92.3 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
 Identities = 140/394 (35%), Positives = 236/394 (59%)

Query: 144 ESKSLNPFQKADSEPEPLNASYVYKEHPFILNLQEDQIENLKQQL-GILVQGQEVTRPI 202
 ++ KL P P ++ Y E P + + + + + + ++ GI V+G+ +PI
 Sbjct: 313 QQRKLLPEVDHGKIEYEPFRKNF-YVEVPELAKMSQEEVNVFRLEMEGITVKGKCPKPI 371
 Query: 203 IDFEHCSLPEVLNHNKKGSGYEVPTPIQMIPVGLLGRDILASADTGSGKTAFFLLPV- 261


```

      + C +   ++LKK GYE PTPIQ Q IP + GRD++ A TGSGKT AFLLP+
Sbjct: 372 KSWVQCISMKILNSLKKHGYEKPTPIQTQAIPAIMSGRDLIGIAKTGSGKTIAFLLPMPF 431
Query: 262 --IM--RALFESKTPSALILTPTRELAIQIEROAKELMSGLPRMKTVLLVGGGLPLPPQLY 317
      IM R+L E + P A+I+TPTRELA+QI ++ K+   L ++ V + GG + Q+
Sbjct: 432 RHIMDQRSLEEGERPIAVIMTPTRELALQITKECKFSKTLG-LRVVCVYGGTGISEQIA 490
Query: 318 RLQOHVKVVIATPGRLLDIIKQSS---VELCGVKIVVVDEADTMLKMGFQQQVLDILENI 374
      L++ ++I+ TPGR++D++ +S   L V VV+DEAD M MGF+ QV+ I++N+
Sbjct: 491 ELKRGAEIIVCTPGRMIDMLAANSGRVTNLRRVTVVVLDEADRMFGFEPQVMRIVDNV 550
Query: 375 PNDCQTILVSATIPTSIEQLASQLLHNPVRIITGEKNLPCANVRQIILWVEDPAKKKKLF 434
      D QT++ SAT P ++E LA ++L P+ + G +++ C++V Q ++ +E+ K KL
Sbjct: 551 RPDQRTVMFSATFPRAMEALARRILSKPIEVQVGGRSVVCSDVEQQVIVIEEEKFLKLL 610
Query: 435 EILNDKKLFKPPVLVFDVCKLGADLLSEAVQKITGLKSIHSEKSQIERKNILKGLLEG 494
      E+L +   V++FVD + AD L + + + +   +S+H Q +R +I+   G
Sbjct: 611 ELLGHYQE-SGSVIIIFVDKQEHADGLLKDLMRAS-YPCMSLHGGIDQYDRDSIINDFKNG 668
Query: 495 DYEVVVSTGVLRGLDLISVRLVVNFDMFSSMDEYVHQ 532
      +++V+T V RGLD+ + LVVN+ P+ ++YVH+
Sbjct: 669 TCKLLVATSVAAARGLDVKHLILVVNYS CPNH YEDYVHR 706

Score = 37 (5.6 bits), Expect = 1.6e-60, Sum P(2) = 1.6e-60
Identities = 13/36 (36%), Positives = 17/36 (47%)

Query: 132 KAKHLLQVKEKEE---KSKLSNPQKADSEPEPLNA 164
      KA++ + KEK E   SK   K D E E   +A
Sbjct: 113 KAENRSRSKEKAEGGDSKEKKKDKDKEDEKEKDA 148

```

Pedant information for DKFZphfbr2_23b10, frame 1

Report for DKFZphfbr2_23b10.1

```

[LENGTH]      580
[MW]           64572.24
[PI]           6.13
[HOMOL]        TREMBL:CEF01F1_1 gene: "F01F1.7"; Caenorhabditis elegans cosmid F01F1. 8e-61

[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-53
[FUNCAT]       04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 5e-53
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-49
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-49
[FUNCAT]       j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 2e-46
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 3e-43
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 4e-39
[FUNCAT]       l genome replication, transcription, recombination and repair [H.
influenzae, HI0892] 3e-35
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 6e-34
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 3e-32
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 8e-30
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 5e-23
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 1e-16
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 5e-11
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 1e-06
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 1e-06
[BLOCKS]       BL00115B Eukaryotic RNA polymerase II heptapeptide repeat proteins
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 6e-53
[PIRKW]        RNA binding 9e-52
[PIRKW]        DEAD box 2e-43
[PIRKW]        transmembrane protein 1e-21
[PIRKW]        DNA binding 5e-48
[PIRKW]        ATP 4e-57
[PIRKW]        purine nucleotide binding 2e-43
[PIRKW]        P-loop 4e-57
[PIRKW]        hydrolase 6e-42
[PIRKW]        protein biosynthesis 2e-43
[PIRKW]        ATP binding 2e-50
[SUPFAM]        WW repeat homology 1e-49
[SUPFAM]        translation initiation factor eIF-4A 2e-43
[SUPFAM]        DEAD/H box helicase homology 4e-57
[SUPFAM]        recQ helicase homology 8e-06

```

```

{SUPFAM}      unassigned DEAD/H box helicases 4e-57
{SUPFAM}      ATP-dependent RNA helicase DBP1 2e-53
{SUPFAM}      ATP-dependent RNA helicase DHH1 6e-40
{SUPFAM}      tobacco ATP-dependent RNA helicase DB10 1e-49
{SUPFAM}      Bloom's Syndrome helicase 8e-06
{PROSITE}     ATP_GTP_A 1
{PROSITE}     LEUCINE_ZIPPER 1
{PROSITE}     MYRISTYL 6
{PROSITE}     CK2_PHOSPHO_SITE 8
{PROSITE}     TYR_PHOSPHO_SITE 1
{PROSITE}     PKC_PHOSPHO_SITE 7
{PROSITE}     ASN_GLYCOSYLATION 1
{PFAM}        Helicases conserved C-terminal domain
{PFAM}        DEAD and DEAH box helicases
{KW}          Alpha_Beta
{KW}          LOW_COMPLEXITY 3.10 %

```

```

SEQ MFVPRSLKIKRNNDDGKSCVAKIIPDPEDLQLDKSRDVPDAVATEAATIDRHISESC
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PFPSPGGQLAEVHSVSEPEQAKDSDHPSEEPVKSFSKTRWAEFGEPIGVVCGRYGEYICD
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ KTDEDVCSLECKAKHLLQVKEKEEKSLSNPQKADSEPEPLNASYVYKEHPFILNLQED
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QIENLKQQLGILVQGEVTRPIIDFEHCSLPEVLNHNKKSQYEVPTPIQMOMIPVGLLG
SEG .....
PRD hhhhhhhhhheeecccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RDILASADTSGSKTAFLLPVIMRALFESKTPSALILTPTRELAIQIERQAKELMSGPLR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ MKTVLLVGGPLPPLQLYRLQHVKVIIATPGRLLDIKQSSVELCGVKIVVDEADTMLK
SEG ...xxxxxxxxxxxxxxxxxxxxx.....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ MGFQQQVLDILENIPNDCQITILVSATIPTSIEQLASQLLHNPVRIITGEKNLPCANVRQI
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ILWVEDPAKKKKLFEILNDKKLFKPPVLVFDCKLGADLLSEAVQKITGLKSISIHSEKS
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QIERKNLKGLEGDYEVVVSTGVLRGLDLISVRLVNFMDPSSMDEYVHQENTYKSTW
SEG .....
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RNPQHQQQDVRMTLGIVGKAQWEEDNQLKVKLGLKKNCS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKF2phfbr2_23b10.1

PS00001	163->167	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	539->542	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	524->528	CK2_PHOSPHO_SITE	PDOC00006
PS00007	489->497	TYR_PHOSPHO_SITE	PDOC00007
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	80->86	MYRISTYL	PDOC00008

PS00008	195->201	MYRISTYL	PDOC00008
PS00008	250->256	MYRISTYL	PDOC00008
PS00008	490->496	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008
PS00017	247->255	ATP_GTP_A	PDOC00017
PS00029	298->320	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphfbr2_23b10.1

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLpPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACQGTSGKTAAAF		
	+LP+ + N+++ G+E PTPIQ+Q IP+ L GRD++A A TGSGKTAAAF		
Query	209	SLPEVLNHNLLKSGYEVPTPIQMMPVGLLGRDILASADTGSGKTAAAF	257
HMM	lIPMLQHIDwdPWpqpPQdPrALILAPTRELAMQIQEEcRkFgkHMNgIR		
	L+P++ + + + ++P ALIL+PTRELA+QI+++++++ + ++ ++		
Query	258	LLPVIMRALFES--KTPS---ALILTPRELAIQIERQAKELMSGLPKMK	302
HMM	ImcIYGGtnMRdQMRmLeRGpPHIVIAATPGRLIDHIERgtldLDrIeMLV		
	++++GG+++ +Q+ +L++ + ++IATPGRL+D+I++ ++ L ++++V		
Query	303	TVLLVGGGLPLPPQLYRLQQHV-KVIIATPGRLDIIKQSSVELCGVKIVV	351
HMM	MDEADRLMDMGFIDQIRrIMrqIPMpwnRQTMMFSATMPdeIqELARrFM		
	DEAD ML MGF+++Q+ +I+ IP + QT++ SAT+P +I++LA ++		
Query	352	VDEADTMLKMGFQQQVLDILENIP--NDCQTILVSATIPTSIEQLASQLL	399
HMM	RNPIRInIdMdElTtnEnIkQwYiyVerEMWKfdClerLle*		
	+NP+RI+ ++++L N++Q++ +VE + K +L+++++		
Query	400	HNPVRIITGEKNLPCA-NVRQIILWVE-DPAKKKKLFEILN	438
HMM_NAME	Helicases conserved C-terminal domain		
HMM	*EileeWLknl.GIrmYIHGdMpQeERdeIMddFnnGEynVLICtdVgg		
	++L+E ++ G++ ++IH+ ++Q ER +I++ +G+Y V ++T V+G		
Query	458	DLLSEAVQKITGLKSISIHSEKSQIERKNILKGLLEGDYEVVSTGVLG	506
HMM	RGIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*		
	RG+D+++V+++V+N+DMP +++ Y++ + T +		
Query	507	RGLDLISVRLVVNFDMPSMDEYVH-QENTYKST	539

DKFZphfbr2_23b21

group: signal transduction

DKFZphfbr2_23b21.1 encodes a novel 193 amino acid protein which is nearly identical to bovine neurocalcin.

Neurocalcin is a Ca(2+)-binding protein with three putative Ca(2+)-binding domains (EF-hands). In cattle, 6 isoforms are differentially expressed in the central nervous system, retina and adrenal gland. Homology with recoverin indicates involvement in Ca2+ dependent activation of guanylate cyclase.

The new protein can find application in modulating/blocking the guanylate cyclase-pathway.

nearly identical to bovine neurocalcin

complete cds complete cDNA
EST hits

Sequenced by AGOWA

Locus: /map="574.6 cR from top of Chr8 linkage group"

Insert length: 3300 bp

Poly A stretch at pos. 3279, polyadenylation signal at pos. 3249

```
1 GGGGAGAATC TGGTGGATGC TGGACCTTGC TGCTGCTGCT ACTGCTGTTT
51 CCAGGGGCTG CAGAGCATGG ACTGTTAAAT CTTGCACTTC TTCTGAGTGA
101 GCTGAATCTT TGCCGCCAGG ATGGGGAAAC AGAACAGCAA GCTGCGCCCG
151 GAGGTCATGC AGGACTTGCT GGAAAGCACA GACTTTACAG AGCATGAGAT
201 CCAGGAATGG TATAAAGGCT TCTTGAGAGA CTGCCCCAGT GGACATTTGT
251 CAATGGAAGA GTTAAAGAAA ATATATGGGA ACTTTTCCCT TTATGGGGAT
301 GCTTCCAAAT TTGCAGAGCA TGCTTCCGCG ACCTTCGATG CAAATGGAGA
351 TGGGACAATA GACTTTAGAG AATTCATCAT CGCCTTGAGT GTAACCTCGA
401 GGGGGGAAGCT GGAGCAGAAG CTGAAATGGG CCTTCAGCAT GTACGACCTG
451 GACGGAAATG GCTATATCAG CAAGGCAGAG ATGCTAGTGA TCGTGCAGGC
501 AATCTATAAG ATGGTTTCCT CTGTAATGAA AATGCCTGAA GATGAGTCAA
551 CCCCAGAGAA AAGAACAGAA AAGATCTTCC GCCAGATGGA CACCAATAGA
601 GACGGAAATC TCTCCCTGGA AGAGTTCATC CGAGGAGCCA AAAGCGACCC
651 GTCCATTGTG CGCCTCCTGC AGTGGGACCC GAGCAGTGCC GGCCAGTTCT
701 GAGCCCTGCG CCCACCAATC GAATTGTAGA GCTGCTTGTC TTCCCTTTTG
751 ATTCTTCTTT TTAACAATTT TTTTTTTTTT TTGCCAAACA ATATCAATGG
801 TGATGCCGTC CCCTGTGCGG TCTGATGCGC CTTCTCCCTG GACGCTTCA
851 GCCTCTTTTG TCGTGGATGC TTCGTGGGAA TGCCCGAGAG CCCAGTGTGC
901 TTGTGGAGAG CATGGACAGA CTTCTGTTGT TTCATTGTTT GATGATTTT
951 AATCGTTACT ATTATTCTT TTTATTCTAA TGTCTCTGTT CTAAACGTA
1001 AGACTCGGGG GTTGGGGCAA AAGAAGGGAA ACCCATCCAG TCCTGTGATT
1051 CTATTGCAAG CTTCAAGGGG CTTTGTGTTG AAAGACAAAA CTCCCCACCT
1101 GGGTCTGTTG TCACACGTGC CGTAGGGGTG ATGGATGGCA CCGGATGCTG
1151 GATTCGCCAA GAACAAGTTA CCCTCTGGGG TGAGGCTATT CCAGCGAGCT
1201 GGGACATTTT CCATGGGGG CCCACTCCCC TCTCTTCCCC AGCAGCTGTG
1251 AGTTTCTAAG CTGTGAACAT TTCAAGATAA ATTAACAGAG GAGAGGAAAA
1301 AGATGGCTCA GCTATTTTTT CACAGGTTTA CACTAGTTGA GCTAATATGC
1351 GTGCTTTTGG AAATTAACA CAAATGGTAA CATATTCCAA AACCAGACCC
1401 ATCTTGTGTC CTATTGTGAT AAAATAAAAA GACGGCTGTA TATAACATAT
1451 TGGGTAAATGC AGACCAAATT AAGTGTTTTG CTTGTTTAA ATGAAATGCA
1501 TGTTTAGTGA GCACTAATAC AATCTTATTC CAGAAGACTG TTTTGTAGT
1551 CTTATTGTGA AGTAAGACAA CTATAATGAA TGTCTGTCTT GTTTGGAAGT
1601 CATATCTGTC TTGACACAAA TGTACCAATC GACAAGTATA TTTTATATAT
1651 TCCATAAAAA TACAAAGTAA CCCTGACTAG GGCCCAACTT TAATTTTGAA
1701 TGCATTTCCA GAGTGGCCAT GCCTAGAGGG CAGATGCAGA GCAGGTGGTA
1751 GTGGGACAGG ACAATTGGAG CACAGGAATG TTAACATGTA TGACAGGGGA
1801 CCAGTAGGGT GGTTCCTCTC TCAGGCCCAG CAGCCCATTG ACAGCATTAG
1851 ACTGGCGGCA TGGTGCTTTT CTGAGCAGAT CAATACTCTG CAGACTCGAA
1901 AAAACATCAC ATACATTCTT GGAACCTTCC AGTGGTTTAA TCTATGTGCA
1951 TGGTTAGGGA GCCAGGCTG GAATATTCAG TTTCCCTGCC CCTGTTAAAG
2001 AATCAGAGGT TGGCAGTCA TCAAAATCAT CATAAAGACA TGGGCAAGTG
2051 TGTCTGTGGT TTCCAAGGCC CCCCTATGGA GAATCCAAAA GTATTTTCCA
2101 TTGCGGTGCT CTTTGAATGC AGACTTCTAT TTCCAGAAAT GACAGCACAA
2151 GTCTGAGTTG CTGTTTGGTC TGGTGACCTC AGACACACTA ATTTGAATTG
2201 AAAGCTAAGA GTAAAAATTT GCTGGTTACA GGCGAGTCAT ACTCTTGCAA
2251 GTAGTTAGCA AAGGGAGGCC CAAATTCTCA AGGTTGTTGA TGGGGAACCT
2301 GCCACTAAGA GAAGGCAGAG AGGTCCCTAG TGGGTATATT TGCTGCCAAG
2351 CCACTTGCCA AAGAAGAGGA ACCACAGAAA GAGAGACATC ATGACCAGGA
2401 GAAAAATGTG ACTAGACATG CTAACCTCCA GGTTTTATA TATGACTTGA
2451 GTCTGCTGTA ATTGGCAGCA GAAATCCAAA TTTGTATGGT AGACCAAAAA
2501 GAACCAATC CATAGGGTGA AATTTTGAGA CCTAGACTCT GTAAAAATAA
```

```

2551 TCCTAGTCTT CCTCCAGGGG TCAGTTCCTC ACAGTGGTTC TGTACCAAAA
2601 CTGCCCCAAT TCCTCCATGG CCAAGTGTTA AAATCTGTGT TTGGAAAAATA
2651 GCGAATTAACTAAGACACA GAAGGCAGAC TGGGTGAGGA GACCTAGCAT
2701 GCCCTATTGG CAGTGCTCAG GAGCTGCATC CCACTTTTCC CTGCTCTGAA
2751 TCGAAGTCCT AGTTCCTTCC TTTGATTCTC CTTGGTAGG TGGAAATCAGT
2801 TAATGTTTTG AGAAACCTGC CTGGGCTCTG CCCTTAGTCA TGACATCTCG
2851 CTGAGCCAGA CCCACTCTGT TCCTTGGAAC CTAGAGCTGG AGTGAGGAGT
2901 AGAGGTCTCC GGCTATTCCA GAAAGAAAAG TGAGCCACAT GCAGGCTGAT
2951 GAATGCCGAC ACTTCCAGAA TGTATAGAAA TAGTCCCTGT CCTGGCCTGC
3001 CACTGACCCT GTCTGTATT TCTCGGAGGT TGTTTTCTC CTTCTCCTTC
3051 CCAGGAAGGT CTTTGTATGT CGAATCCAGT GCACTCAAAGT TTGGCCAAGG
3101 GACTCCACAG CACCCAGAGG ACTGCATGCC TCAAGGTTTA TGCTACTCCT
3151 CTGCTGGGCT GTTCATTGTC ATTGCTGTGT TCAGGGACCT TTGGAATAAA
3201 AACCTTTTCT GTCCCAAATA AAACCAGCCT GTGATGTTCA AGGGACTGGA
3251 ATAAAGTGGC TTACGACCTG AAGGATTCTA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HS431350 from database EMBL:
human STS WI-15914.
Score = 1308, P = 3.1e-53, identities = 276/285

Entry HSG19929 from database EMBL:
human STS A002C26.
Score = 926, P = 1.5e-35, identities = 186/187

Entry AF052142 from database EMBL:
Homo sapiens clone 24665 mRNA sequence.
Score = 7378, P = 0.0e+00, identities = 1482/1487
3' UTR

Medline entries

93247712:
Neurocalcin family: a novel calcium-binding protein abundant in bovine central nervous system.

94045365:
Distinct regional localization of neurocalcin, a Ca(2+)-binding protein, in the bovine adrenal gland.

96407688:
Crystallization and preliminary X-ray crystallographic studies of recombinant bovine neurocalcin delta.

96066284:
Distribution pattern of three neural calcium-binding proteins (NCS-1, VILIP and recoverin) in chicken, bovine and rat retina.

Peptide information for frame 1

ORF from 121 bp to 699 bp; peptide length: 193
Category: strong similarity to known protein
Prosites motifs: EF_HAND (73-86)
EF_HAND (109-122)
EF_HAND (157-170)

```

1 MGKQNSKLRP EVMQDLLEST DFTEHEIQEW YKGFLRDCPS GHLSMEEFKK
51 IYGNFFPYGD ASKFAEHVFR TFDANGDGTI DFREFIIALS VTSRGKLEQK
101 LKWFASMYDL DGNGYISKAE MLVIVQAIYK MVSSVMKMPD DESTPEKRTE
151 KIFRQMDTNR DGKLSLEEFI RGAKSDFPSIV RLLQCDPSSA GQF

```

BLASTP hits

Entry JH0616 from database PIR:
neurocalcin (clone pCalN) - bovine

Score = 1001, P = 5.2e-101, identities = 192/193, positives = 192/193

Entry GGU91630_1 from database TREMBL:
product: "neurocalcin"; Gallus gallus neurocalcin mRNA, complete cds.
Score = 998, P = 1.1e-100, identities = 191/193, positives = 192/193

Entry NECD_BOVIN from database SWISSPROT:
NEUROCALCIN DELTA.
Score = 996, P = 1.8e-100, identities = 191/192, positives = 191/192

Entry S47565 from database PIR:
BDR-1 protein - human
Score = 934, P = 6.6e-94, identities = 174/193, positives = 187/193

Entry I50676 from database PIR:
gene Rem-1 protein - chicken >TREMBL:GGREM1_1 gene: "Rem-1"; G.gallus
rem-1 mRNA
Score = 933, P = 8.4e-94, identities = 174/193, positives = 186/193

Alert BLASTP hits for DKFZphfbr2_23b21, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23b21, frame 1

Report for DKFZphfbr2_23b21.1

```
[LENGTH]      193
[MW]           22215.30
[pI]           5.35
[HOMOL]        PIR:JH0616 neurocalcin (clone pCalN) - bovine 1e-109
[FUNCAT]        98 classification not yet clear-cut [S. cerevisiae, YDR373w] 3e-54
[FUNCAT]        30.03 organization of cytoplasm [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]        03.07 pheromone response, mating-type determination, sex-specific proteins
                  [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]        03.01 cell growth [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]        13.04 homeostasis of other ions [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]        04.05.01.04 transcriptional control [S. cerevisiae, YKL190w] 2e-18
[FUNCAT]        30.04 organization of cytoskeleton [S. cerevisiae, YBR109c] 0.001
[FUNCAT]        08.19 cellular import [S. cerevisiae, YBR109c] 0.001
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YBR109c] 0.001
[FUNCAT]        03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR109c]
0.001
[FUNCAT]        10.02.99 other morphogenetic activities [S. cerevisiae, YBR109c] 0.001
[FUNCAT]        30.05 organization of centrosome [S. cerevisiae, YBR109c] 0.001
[BLOCKS]       BL00018
[SCOP]          direc_ 1.34.1.5.18 Recoverin [bovine (Bos taurus) 8e-55
[SCOP]          dijsa_ 1.34.1.5.17 Recoverin [human (Homo sapiens) 5e-58
[SCOP]          ditcob_ 1.34.1.5.16 Calcineurin regulatory subunit (B-chain 1e-06
[SCOP]          d2mysc_ 1.34.1.5.15 Myosin Regulatory Chain [chicken (Gallu 2e-29
[SCOP]          discmc_ 1.34.1.5.14 Myosin Regulatory Chain [bay scallo 5e-33
[SCOP]          d2mysb_ 1.34.1.5.13 Myosin Essential Chain [chicken (Gallu 4e-26
[SCOP]          discmb_ 1.34.1.5.12 Myosin Essential Chain [bay scallo 6e-27
[SCOP]          diclm_ 1.34.1.5.11 Calmodulin [Paramecium tetraurelia 1e-15
[SCOP]          d4cln_ 1.34.1.5.10 Calmodulin [Drosophila melanogaster 2e-16
[SCOP]          dicfc_ 1.34.1.5.9 Calmodulin [African frog (Xenopus laevis) 2e-16
[SCOP]          dlahr_ 1.34.1.5.8 Calmodulin [chicken gallus gallus 4e-16
[SCOP]          d3cln_ 1.34.1.5.7 Calmodulin [rat (Rattus rattus) 2e-16
[SCOP]          dltrcb_ 1.34.1.5.6 Calmodulin [bovine (Bos taurus) 8e-08
[SCOP]          dicll_ 1.34.1.5.5 Calmodulin [human (Homo sapiens) 2e-16
[SCOP]          dlrtpl_ 1.34.1.4.5 Parvalbumin [rat (Rattus rattus) 8e-06
[SCOP]          d5tnc_ 1.34.1.5.2 Troponin C [turkey (Meleagris gallopavo) 3e-13
[SCOP]          dlpvaa_ 1.34.1.4.3 Parvalbumin [pike (Esox lucius) 6e-06
[SCOP]          dltnp_ 1.34.1.5.1 Troponin C [chicken (Gallus gallus) 9e-11
[EC]            2.7.1.107 Diacylglycerol kinase 2e-08
[PIRKW]         blocked amino end 1e-100
[PIRKW]         phosphotransferase 2e-08
[PIRKW]         duplication 4e-17
[PIRKW]         tandem repeat 7e-06
[PIRKW]         heterodimer 4e-17
[PIRKW]         heart 6e-09
[PIRKW]         zinc 2e-08
[PIRKW]         serine/threonine-specific protein kinase 1e-06
[PIRKW]         muscle contraction 1e-08
[PIRKW]         acetylated amino end 4e-09
[PIRKW]         ATP 2e-08
[PIRKW]         skeletal muscle 6e-09
```

```

[PIRKW]      signal transduction 1e-91
[PIRKW]      protein kinase 2e-08
[PIRKW]      calcium binding 1e-100
[PIRKW]      alternative splicing 2e-13
[PIRKW]      methylated amino acid 1e-09
[PIRKW]      thin filaments 1e-08
[PIRKW]      lipoprotein 1e-101
[PIRKW]      cardiac muscle 6e-09
[PIRKW]      muscle 6e-09
[PIRKW]      myristylation 1e-100
[PIRKW]      EF hand 1e-101
[PIRKW]      retina 2e-51
[SUPFAM]     calcium-dependent protein kinase 2e-08
[SUPFAM]     unassigned calmodulin-related proteins 8e-41
[SUPFAM]     spec-related protein LpS1 7e-06
[SUPFAM]     calmodulin repeat homology 1e-101
[SUPFAM]     human diacylglycerol kinase 2e-08
[SUPFAM]     protein kinase C zinc-binding repeat homology 2e-08
[SUPFAM]     protein kinase homology 2e-08
[SUPFAM]     calmodulin 1e-101
[PROSITE]    EF_HAND 3
[PROSITE]    CK2_PHOSPHO_SITE      7
[PROSITE]    PKC_PHOSPHO_SITE      3
[PFAM]       EF_HAND
[KW]         All_Alpha
[KW]         3D

SEQ      MGKQNSKLRPEVMQDLLESTDFTEHEIQEWYKGFRLDCPSGHLSMEEFKKIYGNFFPYGD
1rec-    .....HHHHHHHHHTTTCCCHHHHHHHHHHHHHHTTTTTEEEHHHHHHHHHHHTTTTC

SEQ      ASKFAEHVFRFTDANGDGTIDFREFIIALSVTSRGKLEQKLKWAFSMYDLGNGYISKAE
1rec-    HHHHHHHHHHHH-----CEEEHHHHHHHHHHHHHCCCGGGHHHHHHHHHTTTTCCCEEHHH

SEQ      MLVIVQAIYKMVSSVMKMPDESTPEKRTTEKIFRQMDTNRDGKLSLEEFIRGAKSDPSIV
1rec-    HHHHHHHHHHHHCCCTTGGGCTTTTCHHHHHHHHHHHHCCCTTTTEECHHHHHHHHHHCHHHH

SEQ      RLLQCDPSSAGQF
1rec-    HHHCCCH.....

```

Prosites for DKFZphfbr2_23b21.1

PS000005	92->95	PKC_PHOSPHO_SITE	PDOC00005
PS000005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS000005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS000006	23->27	CK2_PHOSPHO_SITE	PDOC00006
PS000006	44->48	CK2_PHOSPHO_SITE	PDOC00006
PS000006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS000006	117->121	CK2_PHOSPHO_SITE	PDOC00006
PS000006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS000006	158->162	CK2_PHOSPHO_SITE	PDOC00006
PS000006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS000018	73->86	EF_HAND	PDOC00018
PS000018	109->122	EF_HAND	PDOC00018
PS000018	157->170	EF_HAND	PDOC00018

Pfam for DKFZphfbr2_23b21.1

```

HMM_NAME      EF hand

HMM            *MfrMDkDGDGyIDFEFmeMMkem*
               +FR +D +GDG+IDF EF+ +++
Query          68  VFRTFDANGDGTIDFREFIIALSVT      92

30.75   100   128       1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
Query          *EIqEMFrMDkDGDGyIDFEFmeMMkem*
               +++++F+M+D DG+GYI++ E+++++++
dkfzphfbr2    100  KLKWAFSMYDLGNGYISKAEMLVIVQAI      128

Query          176       1   29 dkfzphfbr2_23b21.1 nearly identical to bovine neurocalcin
Alignment to HMM consensus:
HMM            *EIqEMFrMDkDGDGyIDFEFmeMMkem*
               +FR MD+++DG+++ EEF++ K+
Query          148  RTEKIFRQMDTNRDGKLSLEEFIRGAKSD      176

```

DKFZphfbr2_23f2

group: brain derived

DKFZphfbr2_23f2 encodes a novel 182 amino acid protein with weak similarity to *S. pombe* Vps29p.

No informative BLAST results: no predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to Vps29p

complete cDNA, complete cds, EST hits
S. cerevisiae and *S. pombe* Vps29p are involved in vacuolar protein sorting
 part of the cDNA is encoded by HSAC2350, splice pattern 4 exons

Sequenced by AGOWA

Locus: /map="12q24"

Insert length: 1016 bp

Poly A stretch at pos. 996, polyadenylation signal at pos. 974

```

1 GAATGGGGAG GAGCCAGAGG AAGAGGGCGG CGACGGTGGT GGTGACTGAG
51 CGGAGCCCGG TGACAGCATG TTGGTGTGG TATTAGGAGA TCTGCACATC
101 CCACACCGGT GCAACAGTTT GCCAGCTAAA TTCAAAAAAC TCCTGGTGCC
151 AGGAAAAAAT CAGCACATT C TGTGCACAGG AACCTTTGC ACCAAAGAGA
201 GTTATGACTA CCTCAAGACT CTGGCTGGTG ATGTTTCATAT TGTGAGAGGA
251 GACTTCGATG AGAATCTGAA TTATCCAGAA CAGAAAGTTG TGACTGTTGG
301 ACAGTTCAAA ATTGGTCTGA TCCATGGACA TCAAGTTATT CCATGGGGAG
351 ATATGGCCAG CTTAGCCCTG TTGCAGAGGC AATTGTATGT GGACATTCTT
401 ATCTCGGGAC ACACACACAA ATCTGAAGCA TTTGAGCATG AAAATAAATT
451 CTACATTAAT CCAGGTTCTG CCACTGGGGC ATATAATGCC TTGGAAACAA
501 ACATTATTCC ATCATTGTG TTGATGGATA TCCAGGCTTC TACAGTGGTC
551 ACCTATGTGT ATCAGCTAAT TGGAGATGAT GTGAAAGTAG AACGAATCGA
601 ATACAAAAAA CCTTAAAGCC AGGCCTGTCT TGATGATTTT TGGTTTTTTT
651 TCATTGTCCT GTTGAATCA AGTAATTAAT CATTTAAGAG CCACAAAAAT
701 GTATCACTTT TATAATATTT TGCAGTAAA TATAATACCA TCTTCTCTGT
751 TAATACATAA TTGCTCCAAG CTTCTGTGTA ACTATAAGAA TATATTAGT
801 TTACAGTATA TGGATTCTAT GAAAAATGT CCACACACA GTAATTGGTC
851 ACTTGTTAAG AAAAATTTAT CCTTGTAAGT ATCTTCAAAG TTGATATTTG
901 GAACTTTATT CAAAAGTAG TGCATGTGGA GAAAGAATCT AGACTTTCTT
951 GTATACATTT TTCTCTCTC CAGTAATAAA CAATTACCTT TCATTGAAAA
1001 AAAAAA AAAAAA

```

BLAST Results

Entry HSAC2350 from database EMBLNEW:
 Homo sapiens 12q24 PAC P424M6 Length = 167,217

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 68 bp to 613 bp; peptide length: 182
 Category: similarity to known protein
 Prosite motifs: RGD (60-63)

```

1 MLVLVLGDLH IPHRCNSLPA KFKLLVPGK IQHILCTGNL CTKESYDYLK
51 TLAGDVHIVR GDFDENLNYP EQKVVTVGQF KIGLIHGQV IPWGDMSLA
101 LLQRQFDVDI LISGHTHKSE AFEHENKFYI NPGSATGAYN ALETNIIPSF

```


151 VLMDIQASTV VTYVYQLIGD DVKVERIEYK KP

BLASTP hits

Entry CEZK1128_6 from database TREMBL:
"ZK1128.1"; *Caenorhabditis elegans* cosmid ZK1128
Length = 523
Score = 400 (140.8 bits), Expect = 2.3e-37, P = 2.3e-37
Identities = 81/150 (54%), Positives = 106/150 (70%)

Entry S46793 from database PIR:
hypothetical protein YHR012c - yeast (*Saccharomyces cerevisiae*)
Length = 282
Score = 180 (63.4 bits), Expect = 3.7e-37, Sum P(3) = 3.7e-37
Identities = 35/71 (49%), Positives = 44/71 (61%)

Entry AB011824_1 from database TREMBL:
"Vps29"; *Schizosaccharomyces pombe* mRNA for Vps29,
partial cds. *Schizosaccharomyces pombe* (fission yeast)
Length = 176
Score = 189 (66.5 bits), Expect = 2.7e-27, Sum P(2) = 2.7e-27
Identities = 33/72 (45%), Positives = 50/72 (69%)

Alert BLASTP hits for DKFZphfbr2_23f2, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23f2, frame 2

Report for DKFZphfbr2_23f2.2

[LENGTH] 182
[MW] 20445.84
[pI] 6.29
[HOMOL] TREMBL:CEZK1128_6 gene: "ZK1128.8"; *Caenorhabditis elegans* cosmid ZK1128 2e-51
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YHR012w] 1e-27
[FUNCAT] r general function prediction [M. jannaschii, MJ0623] 1e-16
[BLOCKS] BL01269D
[BLOCKS] BL01269A
[PROSITE] RGD 1
[PROSITE] MYRISTYL 4
[PROSITE] PKC_PHOSPHO_SITE 1
[KW] Alpha_Beta

SEQ MLVLVLGDLHIPRCNSLPKFKKLLVPGKIQHILCTGNLCTKESYDYLKTLAGDVHIVR
PRD cccccccccccccccccchhhhhhhhhcEEEEccccccchhhhhhhhhcEEEE

SEQ GDFDENLNYPEQKVVTVGQFKIGLIHGHOVIPWGDMSLALLQRQFDVDILISGHTHKSE
PRD cccccccccccccccccccccccccccccchhhhhhhhhcEEEEcccccccc

SEQ AFEHENKFYINPGSATGAYNALETNIIPSFVLMDIQASTVVTVYVYQLIGDDVKVERIEYK
PRD ccc

SEQ KP
PRD cc

Prosite for DKFZphfbr2_23f2.2

PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	83->89	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	137->143	MYRISTYL	PDOC00008
PS00016	60->63	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_23f2.2)

DKFZphfbr2_23124

group: intracellular transport and trafficking

DKFZphfbr2_23124.2 encodes a novel 348 amino acid protein with similarity to human glycoprotein gp36b and canine VIP36 glycoprotein.

The vesicular protein VIP36 (36 kDa vesicular integral membrane protein) shows homology to leguminous plant lectins. The protein is localized to the Golgi apparatus, endosomal and vesicular structures and the plasma membrane. VIP36 binds to sugar residues of glycosphingolipids and/or glycosylphosphatidyl-inositol anchors and might provide a link between the extracellular/luminal face of glycolipid rafts and the cytoplasmic protein segregation machinery. Gp36 is located within the endoplasmic reticulum. For the novel protein, a lectin character is predicted. Due to the intracellular localisation of the homolog proteins, it should be involved in intracellular transport and trafficking.

The new protein can find application in modulating/blocking intracellular transport and trafficking.

strong similarity to human GP36b glycoprotein

complete cDNA, complete cds, EST hits
potential start at Bp 29 matches kozak consensus ANNatgG
similarity to lectins,

Sequenced by AGOWA

Locus: /map="2"

Insert length: 2416 bp

Poly A stretch at pos. 2394, no polyadenylation signal found

```

1  GGGGGATGAA  GGGTCGTGG  TGGGAAAGAT  GCGGGCGACT  CTGGGACCCC
51  TTGGGTCGTG  GCAGCAGTGG  CGGCGATGTT  TGTGCGCTCG  GGATGGGTCC
101 AGGATGTTAC  TCCTTCTTCT  TTTGTTGGGG  TCTGGGCAGG  GGCCACAGCA
151 AGTCGGGGCG  GGTCAAACGT  TCGAGTACTT  GAAACGGGAG  CACTCGCTGT
201 CGAAGCCCTA  CCAGGGTGTG  GGCACAGGCA  GTTCCTCACT  GTGGAATCTG
251 ATGGGCAATG  CCATGGTGAT  GACCCAGTAT  ATCCGCCTTA  CCCCAGATAT
301 GCAAAGTAAA  CAGGGTGCCCT  TGTGGAACCG  GGTGCCATGT  TTCCTGAGAG
351 ACTGGGAGTT  GCAGGTGCAC  TTCAAATCC  ATGGACAAGG  AAAGAAGAAAT
401 CTGCATGGGG  ATGGCTTGGC  AATCTGTGAC  ACAAGGATC  GGATGCAGCC
451 AGGGCCTGTG  TTTGGAACA  TGGACAAATT  TGTGGGCTCG  GGAGTATTTG
501 TAGACACCTA  CCCCAGTGG  GAGAAGCAGC  AAGAGCGGGT  ATCCCTTAC
551 ATCTCAGCCA  TGGTGAACAA  CGGCTCCCTC  AGCTATGATC  ATGAGCGGGA
601 TGGGCGGCGT  ACAGAGCTGG  GAGGCTGCAC  AGCCATTGTC  CGCAATCTTC
651 ATTACGACAC  CTTCTGGTG  ATTCGCTACG  TCAAGAGGCA  TTTGACGATA
701 ATGATGGATA  TTGATGGCAA  GCATGAGTGG  AGGGACTGCA  TTGAAGTGCC
751 CGGAGTCCGC  CTGCCCCGCG  GCTACTACTT  CGGCACCTCC  TCCATCACTG
801 GGGATCTCTC  AGATAATCAT  GATGTCAATT  CTTGAAGTT  GTTTGAACGT
851 ACATGGGAGA  GAACCCAGA  AGAGGAAAG  CTCCATCGAG  ATGTGTTCTT
901 GCCCTCAGTG  GACAATATGA  AGCTGCCTGA  GATGACAGCT  CCACCTGCCG
951 CCCTGAGTGG  CCTGGCCCTC  TTCTCATCG  TCTTTTCTC  CCTGGTGTTC
1001 TCTGTATTG  CCATAGTCAT  TGGTATCATA  CTCTACAACA  AATGGCAGGA
1051 ACAGAGCCGA  AAGCGCTTCT  ACTGAGCCCT  CTGTGCGCA  CCACTTTTGT
1101 GACTGTACAC  CATGAGGTAT  GGAAGGAGCG  GGCCTGGCC  TGAGCATGCA
1151 GCCTGGAGAG  TGTCTTGTC  TCTAGCAGCT  GGTGGGGAC  TATATTCTGT
1201 CACTGGAGTT  TTGAATGCAG  GGACCCGCA  TTCCCATGGT  TGTGCATGGG
1251 GACATCTAAC  TCTGGTCTGG  GAAGCCACCC  ACCCCAGGGC  AATGCTGTG
1301 TGATGTGCCT  TTCCCTGCAG  TCCTTCCATG  TGGGAGCAGA  GGTGTGAAGA
1351 GAATTTACGT  GGTGTGATG  CCAAATCAC  GGAACAGAA  TTCAATAGCCC
1401 AGGCTGCCGT  GTTGTGAC  TCAGAGGCC  CTCTACTTC  AGTTTGAAT
1451 CCACAAAGAA  TTAATAACTG  GTAACACCAC  AGGCTTCTG  ACCATCCATT
1501 CGTTGGGTTT  TGCATTTGAC  CCAACCTCT  GCCTACCTGA  GGAGCTTTCT
1551 TTGGAACCA  GGATGAAAC  TTCTCCCTG  CCTTACCTTC  CTTTCACTCC
1601 ATTCATTGTC  CTCTCTGTG  GCAACCTGAG  CTGGGAAAGG  CATTGGAATG
1651 CCTCTCTGTT  GGGGCTGGG  GCTGCAGAAC  ACACCTGCGT  TTCGCTGGCC
1701 TTCATTAGGT  GGGCCTAGGG  AGATGGCTTT  CTGCTTTGGA  TCACTGTTCC
1751 CTAGCATGGG  TCTTGGGTCT  ATTGGCATGT  CCATGGCCTT  CCCAATCAAG
1801 TCTCTTCAGG  CCCTCAGTGA  AGTTTGGCTA  AAGGTTGGTG  TAAAAATCAA
1851 GAGAAGCCTG  GAAGACCCA  TGGATGCCAT  GGATTAGCTG  TGCAACTGAC
1901 CAGCTCCAGG  TTTGATCAAA  CCAAAGCAA  CATTGTTCAT  GTGCTGTGAC
1951 CATGTGGAGA  TGTTCCTGGA  CTTGCTAGAG  CCGCTTAGC  TGCATGTTT
2001 GTAGTTACGA  TTTTGGAAAT  CCCTCTTTGA  GTGCTGAAAG  TGTAAGGAAG
2051 CTTTCTTCTT  ACACCTTGGG  CTTGGATATT  GCCCAGAGAA  GAAATTTGGC
2101 TTTTCTTCTT  TAATGGACAA  GGGCAGATTG  CTGTTCTCAT  GTTCCAAGTC
2151 TGAGAGCAAC  AGACCTCAT  CATCTGTGCC  TGAAGAGTT  CACTGTCATT
2201 GAGCAGACA  GCCTGAGTGC  TGGCTCTGT  CAACCTTAT  TCCACTGCCT

```

2251 TATTTGACAA GGGGTTACAT GCTGCTCACC TTACTGCCCT GGGATTAAAT
 2301 CAGTTACAGG CCAGAGTCTC CTTGGAGGGC CTGGAACCTCT GAGTCTCCT
 2351 ATGAACCTCT GTAGCCTAAA TGAAATTCCT AAAATCACCG ATGGAACCAA
 2401 AAAAAAAAAA AAAAAA

BLAST Results

Entry HS622145 from database EMBL:
 human STS WI-6746.

Score = 1079, P = 5.1e-43, identities = 219/223

Entry G42541 from database EMBLNEW:

SHGC-58649 Human Homo sapiens STS genomic, sequence tagged site.
 Score = 1091, P = 1.7e-43, identities = 219/220

Medline entries

94265253:

A putative novel class of animal lectins in the secretory pathway
 homologous to leguminous
 lectins.

94208543:

VIP36, a novel component of glycolipid rafts and exocytic carrier
 vesicles in epithelial cells.

Peptide information for frame 2

ORF from 29 bp to 1072 bp; peptide length: 348
 Category: strong similarity to known protein

1 MAATLGPLGS WQQWRRCLSA RDGSRMLLLL LLLGSGQGFPQ QVGAGQTFEY
 51 LKREHSLSKP YQGVGTGSSS LWNLMGNAMV MTQYIRLTPD MQSKQGLAWN
 101 RVPCFLRDWE LQVHFQIHGO GKKNLHGDGL AIWYTKDRMQ PGPVFGNMDK
 151 FVGLGVFVDT YPNEEKQOER VFPPYISAMVN NGSLSYDHER DGRPTLGGC
 201 TAIVRNLHYD TFLVIRYVKR HLTIMMDIDG KHEWRDCIEV PGVRLPRGY
 251 FGTSSITGDL SDNHQVISLK LFELTVERTP EEEKLHROVF LPSVDNMKLP
 301 EMTAPLPPLS GLALFLIVFF SLVFSVFAIV IGIILYNKWO EQSRKRFP

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23124, frame 2

PIR:G01447 GP36b glycoprotein - human, N = 1, Score = 1001, P = 5.9e-101

SWISSPROT:VP36 CANFA VESICULAR INTEGRAL-MEMBRANE PROTEIN VIP36
 PRECURSOR (VIP36)., N = 1, Score = 990, P = 8.6e-100

TREMBL:CET04G9_2 gene: "T04G9.3"; Caenorhabditis elegans cosmid
 T04G9., N = 1, Score = 614, P = 6e-60

PIR:S42626 ER-golgi intermediate compartment protein - human, N = 2,
 Score = 397, P = 1e-42

>PIR:G01447 GP36b glycoprotein - human
 Length = 356

HSPs:

Score = 1001 (150.2 bits), Expect = 5.9e-101, P = 5.9e-101
 Identities = 197/356 (55%), Positives = 256/356 (71%)

Query: 1 MAATLGPLGSWQQWRRCLSARDG-----SRMLLLLLLLGSGQGPPQVGAGQTFEYLK 52
 MAA G + W RRCL R G + L LLLLGS + G + E+LK
 Sbjct: 1 MAAE-GWIWRNGWGRRCLG-RPGLLGPGPGPTTFLLLLLLGSVTA--DITDGNS-EHLK 55

Query: 53 REHSLSKPYQGVGTGSSSLWNLGMNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQ 112
 REHSL KPYQGVG+ S LW+ G+ M+ +QY+RLTPD +SK+G++WN PCFL+DWE+
 Sbjct: 56 REHSLIKPYQGVGSSSMLWDFQGSTMLTSQYVRLTPDERSKEGSIWNHQPFLKDWEMH 115

Query: 113 VHFKIHGQGGKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFVGLGVFVDITYPNEEKQQERV 172
 VHF+HG GKKNLHGDG+A+WYT+DR+ PGPVFG+ D F GL +F+DTYPN+E ERVF
 Sbjct: 116 VHFVHGTGKKNLHGDGIALWYTRDLVPGPVFGSKDNFHLAIFLDITYPNDETT-ERVF 174

Query: 173 PYISAMVNNGSLSYDHERDGRPTLGGCTAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKH 232
 PYIS MVNNGSLSYDH +DGR TEL GCTA RN +DTFL +RY + LT+M D++ K+
 Sbjct: 175 PYISVMVNNGSLSYDHSKDGRTLAGCTADFRNRDHDFTLAVRYSRGRLTVMTDLEDKN 234

Query: 233 EWRDCIEVPGVRLPRGYFYTSSITGDLSNDHVISLKLFEVERTPEEEKLHRDVF 292
 EW++CI++ GVRLP GYYFG S+ TGDLSNDHD+IS+KLF+L VE TP+EE + P
 Sbjct: 235 EWKNCIDITGVRLPTGYFYGASAGTGDLSNDHDIISMKLFQLMVEHTPDEESIDWTKIEP 294

Query: 293 SVDNMKLPENTAPLP-----PLSGLALFLIVFSLVFSVFAIVIGIILYNKWQEQSRK 345
 SV+ +K P+ P PL+G +FL++ +L+ V V+G +++ K QE++ K
 Sbjct: 295 SVNFLKSPKDNVDDPTGNFRSGPLTGWRVFLLLCALLGIVVCAVGVAVFQKRQERN-K 353

Query: 346 RFY 348
 RFY
 Sbjct: 354 RFY 356

Pedant information for DKFzphfbr2_23124, frame 2

Report for DKFzphfbr2_23124.2

[LENGTH] 348
 [MW] 39711.10
 [pI] 8.55
 [HOMOL] PIR:G01447 GP36b glycoprotein - human 1e-101
 [PIRKW] lectin 2e-37
 [PIRKW] transmembrane protein 2e-37
 [PIRKW] endoplasmic reticulum 2e-37
 [PIRKW] Golgi apparatus 2e-37
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 5
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 3
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] Alpha_Beta
 [KW] SIGNAL_PEPTIDE 39
 [KW] LOW_COMPLEXITY 7.76 %

SEQ MAATLGPLGSWQWRCLSDRGSRMLLLLLLLGSGQGPQGVAGQTFEYLRHSLSKP
 SEGxxxxxxx.....
 PRD cccccccccccccccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhccc

SEQ YQGVGTGSSSLWNLGMNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQVHFKIHGQ
 SEG
 PRD cccccccccceccccccccccccccccccccccccccccccccchhhhhhhhhhhheeeccc

SEQ GKKNLHGDGLAIWYTKDRMQPGPVFGNMDKFVGLGVFVDITYPNEEKQQERVFPYISAMVN
 SEG
 PRD cccccccccceeecc

SEQ NGSLSYDHERDGRPTLGGCTAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKHWRDCIEV
 SEG
 PRD ccc

SEQ PGVRLPRGYFYTSSITGDLSNDHVISLKLFEVERTPEEEKLHRDVFPSVDNMKLP
 SEG
 PRD ccc

SEQ EMTAPLPPLSGLALFLIVFSLVFSVFAIVIGIILYNKWQEQSRKRKY
 SEGxxxxxxx.....
 PRD cccccccccchhhccc

Prosites for DKFzphfbr2_23124.2

PS00001	181->185	ASN_GLYCOSYLATION	PDOC00001
PS00002	35->39	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	19->22	PKC_PHOSPHO_SITE	PDOC00005

WO 01/12659

PCT/IB00/01496

PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS00006	279->283	CK2_PHOSPHO_SITE	PDOC00006
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00009	120->124	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_23124.2)

DKF2phfbr2_23n16

group: signal transduction

DKF2phfbr2_23n16.1 encodes a novel 292 amino acid protein with weak similarity to putative phosphatidylinositol-4-phosphate 5-kinase of *Arabidopsis thaliana*.

The novel proteins contains a WW domain which has been originally described as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. This domain is frequently associated with other domains typical for proteins in signal transduction processes. Examples of proteins containing the WW domain are Dystrophin, Utrophin, vertebrate YAP protein (binds the SH3 domain of the Yes oncoprotein), murine NEDD-4 (embryonic development and differentiation of the central nervous system), IQGAP (human GTPase activating protein acting on ras). Therefore the new protein should be involved in intracellular signal transduction.

The new protein can find application in modulating/blocking intracellular signal transduction pathways.

similarity to putative phosphatidylinositol-4-phosphate 5-kinase

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 2936 bp

Poly A stretch at pos. 2916, polyadenylation signal at pos. 2873

```
1 GGGGGCGCTC CCGAGAAAGA GTGAGGGCGC GACGCGCACC AACGGTGGAG
51 GGATGTTTCA GCAGCCCTTG AGAAGGAAGA GGAGGAAGCT GAGGGCCCGC
101 TGAGGGCGCA GGACCTGAGG GAGTCCTACA TCCAGCTCGT CCAGGGTGTG
151 CAGGAGTGGC AGGATGGTTG CATGTACCAG GGGGAGTTTG GGTGGAACAT
201 GAAGCTTGGG TATGGCAAAT TCTCTGGGCC CACAGGCGAG TCATACCATG
251 GGCAGTTTTA CCGGGACCAC TGCCATGGCC TGGGTACCTA CATGTGGCCA
301 GATGGGTCCA GTTTCACGGG CACATTTTAC CTCAGCCACC GAGAAGGCTA
351 CGGCACCATG TACATGAAGA CACGGCTTTT CCAGACTCAC TGCCACAACG
401 ACATTGTCAA CCTTCTCCTG GACTGTGGGG CCGACGTGAA CAAGTGCTCA
451 GATGAGGGTC TCACGGCACT CAGCATGTGT TTCTCTCTCC ACTACCCCGC
501 CCAGTCTCTT AAGCCCAATG TTGCTGAACG GACCATACCT GAGCCCCAGG
551 AACCTCCAAA ATTCCCAAGT GTTCCAATCC TTTCATCATC ATTTATGGAC
601 ACAAACTTGG AGTCTCTGTA CTATGAGGTG AACGTGCCTT CCCAGGGTAG
651 CTATGAGCTG AGGCCACCGC CAGCACCAC GCTCTGCCA CGCGTCTCAG
701 CGAGCCACGA GGGCGGCCAC TTCCAGGACA CCGGGCAGTG TGGGGGGTCC
751 ATAGACCACA GGAGCAGCTC TCTGAAGGGG GACTCCCCGT TGGTGAAGGG
801 CAGCCTTGGC CATGTGAAA GCGGCTTGA GGAGTGTG TGAGACACAG
851 ACCGGGGCAG TCTGTGCAGT GCTGAGACGA AATTTGAGTC CAACTGTGTG
901 GTGTGCGACT TCTCCATCGA GCTCTGCGAG GCCATGCTGG AGAGAAGCGC
951 CCAGTCCAC AGCTTGCTGA AGATGGCCTC GCCCTCACCG TGCACCAGCA
1001 GCTTCGACAA AGGGACCATG CCGAGGATGG CGCTGTCCAT GATCGAGTAG
1051 GTCCCTGGCA CAGCTGGTGG GGGTGGAGGG CCACCATCAG GGTGAATCC
1101 TATGCTCAGC AGACCCACGT CTCTTCCCTG TGCCAGTGGG AGGCGTGTG
1151 TCTGGAGATG TGTGTCTGAA TGTGTGAGCA TCCCTGTGTC GGTGGCTCCA
1201 TGCCATGGCC AGCCCTGTGG GGGTGCACG GTGACGGGCT GTTTCAGTG
1251 CCACCCACAG CCTGTGGGGG TGCCACGGTG ACGGGCTGTT TTCAGTACCA
1301 CGCCAGCCCT GCTTGGCCCT TTGGCACTGG CCTGAAGTGT CTCTGTGGGA
1351 GCCTCAGCAG GGGCCACTGT CAGGGGTCTT ATCCTAGCCA TAGTGCACGT
1401 GAGTGACACC TGCCTGGGCA GCTCTCACAC CCCTGCTGTC CACCCGTGCT
1451 ATACCACTGT GTCTCAAAT GTGGTCTATG CACCCCGGGG GGTCCAAGAC
1501 CCTTTCAGGG AGTCTGTGGG GTCAAATGA TTCTCTGAT AACCTGAGA
1551 CTCTCTTAGC CTCTCTTTG TGTGATGTT GGTGGATGGT ATGAAGACAG
1601 GCGCGTGCAG ACCACAGCC CCCAGCGTGC AGGGCAGCAG TGCCTGGCCT
1651 GCTTGGGGGC ATGGTATTCC TTCACCACGG TGTGCACTTG CCGGGATGCC
1701 TGTCTCACTG AAGAATGCCT TTGACTAAGC AGAAAAGCAA TGACAAATTG
1751 CATTAAATCT TGCTCCTTGC GTACACACCC CTCGAATATT CTGGGTCGGA
1801 AAACATGGGA AGGACACTGA TGTGTGCTG CCACAGACCA AGGCACACCG
1851 CTTCCCGGCA AGAAGCGCTT CCCCCAGGGC CAGAGTAGCA ACAGAAATGCG
1901 GCATCTTCCC AACCTCTGCG CCCATTTTTG ATTGGAAGAA TGACCACTGG
1951 TATGTGGCTG TTAATTTCTC TGAACACAGC CTGCCACTTT AAGGAAAACA
2001 TATGACACTA TTTGTTGCTG GCGAAATTTA CATTTTCAGG TGAATAGCAG
2051 AATTCTGGAC ACTTGCCACC ACCACAAAA CCTTCATAGC TTCCCTTAAC
2101 TTTGAGACAT GGGGTTCAG AGGTTTTTCA CTGAGATGG CGTTAGCAGC
2151 GCAGTTTTGT GATACTGCCT GAAGACATGC CGACAGTGCC CAGATCTCTT
```

```

2201 CTATTGGTGA GCCAGCTTTT CCCACACGGC CAAGTTCTGA TGTTGAACCA
2251 TTGCCAGGTG GGTGAAGATC CATTGACAGT GAGAGGTGGG CCCGTGGGCT
2301 TCAGTGCAGC CAGGCGCAGA AGGCTGGTTC ATGAGTGTCC AGCTCCGCCA
2351 GGTAGCTAGC TCACCACCCC CAGCCTGGGT TCATGTAGTT CAAATAGGAA
2401 GACCACGATG ATCAGAAAGG CTGCTCAAT ACTCCTTCGT CCAGCCGCGT
2451 ACCTGGGGGA GGCTGAATCT CCACTCACTT CCACCAAGGC TGTGCAGAGC
2501 AGATAGGGGA ATCCAGCAAA GGTGAAAAAC AGTGCCATCC TTCTCCCAA
2551 CTGGTTTTGT TTTGTAAAT AACTTTTTGT GACAGTGTTA CTTATTAGTA
2601 ACATGCAGTG GGTTTGTTAT GGTTAACAAG TTGGTGAGCA TTATTGAGAG
2651 GTGAAGCCAG CTGAGCTTCT GGGTTGGGTG GGGACTTGGA GAACTTTTGT
2701 GTCTAGCTAA AGGATTGTAA ATGCACCAAT CAATGCTCAG TGTCTAGCTA
2751 AAGGATTGTA AATGCACCAA TCAGCACTCT GTAAATTTGA CCAATCAGCG
2801 TTCTGTAAAA TGGACCAATC AGTGGTCTGT AAAATGGACC AGTCAGCAGG
2851 ATGTGGGCGG GGCCAAAAAA GGGGAATAAAA GCTGGCCACC GCCAGGCTCC
2901 CCACACGCTT GCAGCGAAAA AAAAAAAAAA AAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 172 bp to 1047 bp; peptide length: 292
 Category: similarity to unknown protein
 Prosite motifs: WW_DOMAIN_1 (19-24)

```

1 MYQGEFGLNM KLGYGKFSWP TGESYHGQFY RDHCHGLGTY MWPDGSSFTG
51 TFYLSHREGY GTMYMKTRLF QTHCHNDIVN LLLDCGADV N KCSDEGLTAL
101 SMCFLHLHYA QSFKNVAER TIPEPQEPK FPVVPILSSS FMDTNLESY
151 YEYVNPVSQS YELRPPPAFL LLPRVSGSHE GGHFQDTGQC GGSIDHRSSS
201 LKGDSPLVKG SLGHVESGLE DVLGDTDRGS LCSAETKFES NLCVCDFSIE
251 LSQAMLESA QSHSLKMAS PSPCTSSFDK GTMRRMALSM IE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23n16, frame 1

TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds., N = 2, Score = 138, P = 1.1e-06

TREMBL:AF019380_1 product: "putative phosphatidylinositol-4-phosphate
 5-kinase"; Arabidopsis thaliana putative
 phosphatidylinositol-4-phosphate 5-kinase mRNA, complete cds., N = 2,
 Score = 138, P = 1.4e-06

PIR:T02098 probable phosphatidylinositol-4-phosphate 5-kinase -
 Arabidopsis thaliana, N = 2, Score = 135, P = 6.7e-06

>TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for
 AtPIP5K1, complete cds.
 Length = 683

HSPs:

Score = 138 (20.7 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
 Identities = 23/61 (37%), Positives = 35/61 (57%)

```

Query:      1 MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTTFYLSHREGY 60
            MY+G++      G GKFSWP+G +Y G+F      G GT+   DG ++ GT+   + G+
Sbjct:     34 MYEGDWKRKASGKGFSGATYEGFEKSGRMEGFGTGTGADGDTYRGTWVADRKHGH 93

Query:      61 G 61
            G
Sbjct:     94 G 94

```

Score = 112 (16.8 bits), Expect = 9.7e-04, Sum P(2) = 9.7e-04
Identities = 19/51 (37%), Positives = 27/51 (52%)

Query: 12 LGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGYGT 62
+G GK+ W G Y G + R G G + WP G+++ G F EG+GT
Sbjct: 22 IGSGLYLWKGCMYEGDWKRGKASGKGSWPSGATYEGEFKSGRMEGFGT 72

Score = 97 (14.6 bits), Expect = 4.4e-02, Sum P(2) = 4.3e-02
Identities = 19/60 (31%), Positives = 32/60 (53%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY 61
Y+GEF G+G F+ G++Y G + D HG G + +G + GT+ ++G G
Sbjct: 58 YEGEFKSGRMEGFGTFTGADGTYRGTWVADRKHGHGQKRYANGDFYEGTWRRLQDGRG 117

Score = 93 (14.0 bits), Expect = 1.2e-01, Sum P(2) = 1.1e-01
Identities = 18/62 (29%), Positives = 34/62 (54%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY 61
Y+G + + K G+G+ + G+ Y G + R+ G G Y+W +G+ +TG + + G G
Sbjct: 81 YRGTWVADRKHGHGQKRYANGDFYEGTWRRLQDGRGRYVWRNGNQYTGEWRIGVISGK 140

Query: 62 TM 63
+
Sbjct: 141 LL 142

Score = 91 (13.7 bits), Expect = 2.0e-01, Sum P(2) = 1.8e-01
Identities = 18/51 (35%), Positives = 24/51 (47%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY 52
Y GE+ + + G G WP G Y G + G G + W DGSS G +
Sbjct: 127 YEGWRIGVISGKGLLVWPNNGNRYEGLWENGIPKGNVFTWSDGSSCVGAW 177

Score = 90 (13.5 bits), Expect = 2.6e-01, Sum P(2) = 2.3e-01
Identities = 17/60 (28%), Positives = 31/60 (51%)

Query: 2 YQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY 61
Y+G + N++ G G++ W G Y G++ G G +WP+G+ + G + +G G
Sbjct: 104 YEGTWRRLQDGRGRYVWRNGNQYTGEWRIGVISGKGLLVWPNNGNRYEGLWENGIPKNG 163

Score = 45 (6.8 bits), Expect = 1.1e-06, Sum P(2) = 1.1e-06
Identities = 14/62 (22%), Positives = 26/62 (41%)

Query: 215 VESGLEDLVGDTRGSLCSAETKFESNLCVDCF--SIELSQAMLESAQSHSLKMASPS 272
V+SG + G+ +C E+ E+ CD ++E S +R + + +
Sbjct: 205 VDSGAGSLGGEKVFPRICIWESDGEAGDITCDIIDNVEASMIYRDRISVDRDGRQFQKN 264

Query: 273 PC 274
PC
Sbjct: 265 PC 266

Pedant information for DKFZphfbr2_23n16, frame 1

Report for DKFZphfbr2_23n16.1

[LENGTH] 292
[MW] 32214.44
[pI] 5.51
[HOMOL] TREMBL:AB005902_1 product: "AtPIP5K1"; Arabidopsis thaliana mRNA for AtPIP5K1,
complete cds. 7e-08
[BLOCKS] BL01137A Hypothetical YBL055c/yjjv family proteins
[PROSITE] WW_DOMAIN_1 1
[PROSITE] MYRISTYL 5
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] PKC_PHOSPHO_SITE 5
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 4.11 %

SEQ MYQGEFGLNMKLGYGKFSWPTGESYHGQFYRDHCHGLGTYMWPDGSSFTGTFFYLSHREGY
SEG
PRD ccc

SEQ GTMYMKTRLFQTHCHNDIVNLLDCGADVKNKCSDEGLTALSMCFLLHYPAQSFKPNVAER
SEG
PRD cccchhhhhheeccccchhhhhccccccccccccccccchhhhhhhhhccccccccccccc

SEQ TIPEQEPKFPVVPILSSSFMDTNLESLEYEVNVPSQGSYELRPPAPLLLPVSGSHE

WO 01/12659

PCT/IB00/01496

```

SEG .....XXXXXXXXX.....
PRD ecccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GGHFQDTGQC GGSIDHRSSSLKGDSP LVKGS LGHVESGLE DVLGDTDRGSLCSAETKFES
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ NLCVCDFSIELSQAM LERSAQSHSLKMASPSPCTSSFDKGTMRMRMALSME
SEG .....
PRD cccccchhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhccc

```

Prosite for DKFZphfbr2_23n16.1

PS00005	55->58	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	200->203	PKC_PHOSPHO_SITE	PDOC00005
PS00005	226->229	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS00006	121->125	CK2_PHOSPHO_SITE	PDOC00006
PS00006	140->144	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00008	45->51	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	177->183	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS01159	19->44	WW_DOMAIN_1	PDOC50020

(No Pfam data available for DKFZphfbr2_23n16.1)

DKFZphfbr2_23o24

group: brain derived

DKFZphfbr2_23o24 encodes a novel 139 amino acid protein with similarity to CAAX-box proteins.

The CAAX box is a prenyl group binding site found in a number of eukaryotic proteins, such as which is found in Ras- and ras-like proteins such as Rho, Rab, Rac, Ral, and Rap, as well as in nuclear lamins A and B, some G protein alpha and gamma subunits and some dnaJ-like proteins. These proteins are posttranslationally modified at this site by the attachment of either a farnesyl or a geranyl-geranyl group to a cysteine residue.

No informative BLAST results; no predictive prosite, pfam or SCOP motife

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to lectins

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3564 bp

Poly A stretch at pos. 3541, no polyadenylation signal found

```
1 GAATGGCTCC GCAGATGGCC GGCAGTGA GAAGCAAGA AGCGGAGGAG
51 ATGGGCTTC AGCAGGGGGT TGCGGGGGGA GCTTTAACT GAGCCCTGTA
101 AACATGGCAG AACTGCTCAG TGGGAGACTC TCAGCACAGA CGGTCATGGG
151 GAAGTGAGTG CAGTTCATTG GTAATCTTGT TGTCGAGTTC TGGGTTTTTT
201 TTGTTGTTT CGTAACTTTA AAGGTATGCA CTTTATATAG ATTATTTAT
251 TTGCTGGGAC CGTTACTCAG AGTTCCTAGA AATGTACACA GCTTTTTTAC
301 CAGGTTTACT CCTCAGAATC ACTTGTCAC TCTTTAAATG AATGAATGAA
351 TGTGCCAGGC CCTATGCCTG GAGGTTGGGA GCTTCATCTA CATCACATTC
401 TAACAGGTGA CCACTGGGGT AAGCACTGTG TGACTGCAAA GCCAGGGTGT
451 GTTTCATCAT ACACCCAGAT GACCGTGCCT ATGTGCCCTT GTTGCTCTCC
501 CTCCAGGACT GCCTCCTCAC CCCACCCCTT TCTGCAGCTC CTCATCTAAA
551 CATCTCGCCT GGTGAGGTCA CGGCTTAGCC TGTGCGCAG TGCGCCCAAC
601 ACCATCCTTC CCCCTGTGCA GATTGGAGGA GGCCAGGTCT CTCCTCTTAG
651 CTCCTATGTC CCCTTACACC CCCATGGCAC AGATGAGACA TTCACAGAGT
701 TTGCAGATGA TGGAGAGAGG GACTCCAGGT TGCCAGGTGT GTCCACTCTC
751 AGGAACCCCC AGCCCAAGCC TCACTGCTCG TGTTCACAGC CAACCCACAGC
801 ACGGGGGGATA CGCCGGTGCT GTTTCCTGCG TCAGATACAA CCAATTACCA
851 GAAACGACCT CACCCCTCCA ACCACTTTCC AAGGTGCCAG GACAGAGAAG
901 CCCTTCACTG GCCCACCAG GGCAGTTGAC AGAGGGATGC CCTCTTTGGA
951 GGGGAGCCTC ACCTCTACCC ACAGGGCCCG GGCCTTGTC TGGATTCTCA
1001 CCGGGGAGCT CAGCTCAGGA TGGAGAGGTC CCATGTCAGC CAGTTCCTTG
1051 GTGGGGGTCA TGTAGTCTGA AATGACCTGC CGATGGTCCA GGCTGAGCCA
1101 GGGGAGCTGA GCCTGGGTGC CTTTGGGTG CCTACTCTGA CTTGAGTTGG
1151 ATTCATGCCA CAGACCCACC TTCTTGAGCA ACAACACATA TAGCCACCAA
1201 CACAAGAGCC AGGCACACAC TGAGCAGAGA AAGTCCCTGT CGCCTCACCA
1251 CCCAAAAAAT CCAGCTTTGC AGAGACCAAG GTTCTCTCT ACCTTTGAG
1301 AAGCCTCTGT GACCAAAACC GGAGCTTGCC CTCTGAGGC CTCTAGCATT
1351 TCTCCAGGTG TTTTTCAGAG GACTTGGTTT AAATTTGTTT ACCCCAAATG
1401 TGGTCTTTCC CGGATCATGA AAGGATCTGC CGCAAGGTG AATCTGAGTC
1451 TCCTCAGAGT CATATGAGAC TGAACATGCT TATAACATTT CCGTGACCTA
1501 ATAAAGTCTT CAAAAATGTA GGGTATTAA AGTTTAGTGA CATTAAAAAG
1551 TTTAGTCGAA AATATCGTGA TTCAGGTATA TTTAGACATT TGATTCATGC
1601 CAAATTGCCA CTGTTAACAG AAAACACACC CCAAGCACAT TAATGCCTAG
1651 ATATTTCAAA CCCTTTTCTG CCCACACATT CTTAAAAATA ATATACTGAG
1701 AAATCTATAT ACAGGTTTTT TTTAATTAG CTTGGAAAAG AGCAGTTGTA
1751 TTCTGTTTGA ACAGCTGCTA ATGTCAATTC CTGTGGGAAG AAAGACCAAA
1801 GAACATGGAG TTACACCAAG AATTTTAAAA CAAAGACGCT GTCCCTTTCC
1851 TGAGCACCGT GCAGCCAAGA CTGAGAGATC AGTCTGAGAC CTGTGATTAA
1901 GGAGTGTTTT CTACATAGCG TATAATTATG GAGCCACACA AGTGGGCCAT
1951 TACTCTGTTG AGTGCTTCAT GTTTGAGGTA TTTTCGTGTT CCAACTTACA
2001 TTTAAAGTGT TATAAACAG GAAAAATCCA CGAGCAGGTA TTGACACTAT
2051 CCATATTAGA TCATCACAAA ATTATATATA TAGCAGAGTC ATAAACAATG
2101 AGAAACGGCT TTCCCACT TGCTTTAAAT GGCCATGACC TAGTGTTTAG
2151 GGAAGACAGT AAAATCAGCG AGGAGCTCGT GGGAAAAATG AGACGGGCCC
2201 TGAGGGGGTG ACTCATGGGC CAAGCAGGGC CACACAGGTA CCAGGCGGCC
2251 ACGTCTCTCT CTGCTCTCA CTCTCTGGAG ACTGGACTTC CTTTACTGCC
2301 TCCTTTCTGA CATTTCTAG ACATCAGACT TTGCTACTTA GTACACAAAC
2351 GGGGTTCCTT TTTAAATTG TCACTCTAG TTAGCATTTG CAGAAGCTGT
2401 GAAAAATTAC AGAGAGATGA TGTGTTGGGT AAGAGATGGT TTTAAAGTCC
```

```

2451 AGCTTGCTGT TTTTCATTAA GTGCTCTGAA AATGAGTAAG TGGCGTTCCT
2501 GGAGGGGAAC AATCATATAA TTCCGCAGGG TGGGCTCTAA CTTGTTTTCT
2551 GATAGTGTIT AGCAGCTCAT GGCTCTGAGG GCACCTGATA ACACAGCAGC
2601 CAGGCGCTGA TGAGAAAGTG GTGCCAGACA GACCCGAGTG TGGCTTGGCT
2651 CTTGCCCTTAT GTTCCTTTCT CTGTTGAGAG AAGCGTGAGA TGAGATTTTG
2701 TGATTATATT GCACCTCTTG GGCTGACTTT CCCATGCACA GAATGTTTTA
2751 CACATCCTGA TAGCTGAGCT GAAAATGCAA AGAGAAGGGA AAATGCCTTA
2801 AATGTTCTCT GCTAATTTAG AAGCAGCAGG CCTTGGAGT CTTTGTCTCTG
2851 TGTCCCTGAA CAAATCTTAT GGGAGCTCTG GTACCTATGC CAGAAAATGC
2901 ACATAGGCAC AACACTTTTA CATAACGTT CACACACCCC ACCCTTATGG
2951 AGAACTTTT TCTAAATAAG AGAAAGAAAA ATTTTAAGAC TTACAAGTTA
3001 TGTTTAGGTA TTTTACATGG TTCAGAAAAA AAGACATGAA GCGGTATAAA
3051 CTGAGAAGTC TTGTTCCAC AACCCACGT GCCAGGTACA CATAACCATT
3101 TTTATTCACT TCTAGCTTGT GCTTCCAATG TTTGTTAGGC ATATGTAAAT
3151 AAGTGAATAG ATAAGCATTT CTCCTCCTT TTGCTGACAT GAGTGGTGGC
3201 ATGTTTTTGC CCTGGCTTTT ATCCCTTGAC CCCATTCCAG TACCTAGAGA
3251 CCTGCTTCAT TTTTTTAGAT GTGTAATACT TCATGTGTGC GTGTGCCCTTA
3301 GTGATTAAC TGTGCACTGT GCAGGGACAT CGGGCTGGGA TCAGTTTGT
3351 CACTGATATA TACAGCGCTG CGGGAGATAC CCTCACATGT GTATCATTG
3401 GTCCATGTGC AGGTGTGTCT GGAAGATAGA ATTCTAGGCG TAGAATTGAT
3451 AGGTTAAATG TATTTATAGG GAAAAAATCA ATATAAACT TTGCGTGTA
3501 TGATATTGCT GTGCTTTTTT TTTTAATTT TTTACCCAAA TAGTAAAAAA
3551 AAAAAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 656 bp to 1072 bp; peptide length: 139
 Category: similarity to known protein

```

1 MSPSPMAQM RSHSLOMME EKTPGCQVCP LSGTPSPSLT ARVPSQPQHG
51 GYAGAVSLLR YNQLPETTSF LQPLSKVPGQ RSPSLAHPGQ LTEGCPFWRG
101 ASPLPTGPRP CPGFSPGQSR QDGEVPCQPV LWWGSCSLK

```

BLASTP hits

Entry CEEGAP7_1 from database TREMBL:
 gene: "EGAP7.1"; Caenorhabditis elegans cosmid EGAP7.
 Score = 123, P = 2.3e-07, identities = 35/103, positives = 44/103

Entry MMBPC35_1 from database TREMBL:
 Mouse carbohydrate binding protein 35 mRNA, 3' end.
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Entry A28651 from database PIR:
 galactose-specific lectin - mouse >TREMBL:MMMAC2A_1 Mouse mRNA for
 Mac-2 antigen
 Score = 113, P = 2.2e-06, identities = 40/103, positives = 44/103

Alert BLASTP hits for DKFZphfbr2_23o24, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_23o24, frame 2

Report for DKFZphfbr2_23o24.2

```

[LENGTH]      139
[MW]           14748.91
[pI]           8.90
[PROSITE]     PRENYLATION    1

```

```

[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 1
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] PKC_PHOSPHO_SITE 1
[KW] All_Alpha
  
```

```

SEQ MSPSPMAQMRHSQSLQMMEKTPGCQVCPLSGTPSPSLTARVPSQPQHGGYAGAVSLLR
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ YNQLPETTSPLQPLSKVPGQRSPLAHPGQLTEGCPPWRGASPLPTGPRPCPGFSPGQSR
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ QDGEVPCQPVLWWGSCSLK
PRD ccccccccccccccccccccc
  
```

Prosites for DKFZphfbr2_23o24.2

PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00006	119->123	CK2_PHOSPHO_SITE	PDOC00006
PS00008	50->56	MYRISTYL	PDOC00008
PS00013	126->137	PROKAR_LIPOPROTEIN	PDOC00013
PS00294	136->140	PRENYLATION	PDOC00266

(No Pfam data available for DKFZphfbr2_23o24.2)

DKF2phfbr2_23o5

group: brain derived

DKF2phfbr2_23o5 encodes a novel 360 amino acid protein with no known similarity

No informative BLAST results; no predictive prosite, pfam or SCOP motifs

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

potential start at Bp 24 matches Kozak consensus ANNatgG

Sequenced by AGOWA

Locus: /map="7q21-q22"

Insert length: 1736 bp

Poly A stretch at pos. 1714, polyadenylation signal at pos. 1680

```
1 GGGGGAGGAT CAAAGTAGGC AAGATGGCGT CGAGCGCGCG GGAGCCAGGG
51 AGTTTATTGG ATCACCACGT CCAGAGGGCG GTATGCGACA CACGGGCCAA
101 ATATCGAGAG GGACGACGGC CTCGTGCTGT GAAGGTATAT ACAATCAATT
151 TGGAACTCTA GTACTTATTA ATACAAGGAG TTCCTGCTGT GGGAGTCATG
201 AAGGAATTAG TTGAGCGATT CGCTTTATAT GGTGCAATTG AACAGTACAA
251 TGCTCTAGAT GAATACCCAG CAGAAGACTT TACTGAAGTT TATCTTATTA
301 AATTTATGAA CTTACAAAGT GCAAGGACAG CCAAGAGAAA AATGGATGAA
351 CAGAGTTTCT TCGGTGGATT GCTTCATGTG TGCTATGCTC CAGAATTTGA
401 AACAGTTGAA GAAACTAGAA AAAAATACA AATGCGGAAG GCATATGTAG
451 TAAAACTAC TGAAATATA GACCATTACG TGACAAAGAA GAAATTGGTT
501 ACAGAGCATA AAGACACAGA GGATTTTAGA CAAGACTTCC ACTCAGAGAT
551 GTCTGGATTT TGTAAAGCTG CTTTGAACAC TTCTGCAGGG AACTCAAATC
601 CTTATCTTCC GTATTCCTGT GAATTCCTT TATGTTATTT CTCCTCAAAA
651 TGTATGTGTT CATCGGGGGG ACCTGTAGAC AGAGCACCAG ACTCCTCTAA
701 GGATGGTAGA AACCATCATA AAACAATGGG GCATTATAAC CACAATGACT
751 CTTTGCGGAA AACACAGATA AACTCTTTGA AAAACTCAGT GGCCTGCCCT
801 GGTGCACAAA AGGCTATTAC GTCTTCAGAG GCAGTTGACA GATTATGCC
851 TAGGACACAA CACTGCAGG AGCGCAAAAG AAGAAGAGAA GATGATCGTA
901 AACTTGAAC TTTTCTTCAA ACAAAACCAA CTGGTAATGA GATTATGATT
951 GGACCTCTGT TACCAGACAT CTCATAAGTG GATATGCACG ATGACTCATT
1001 GAATACAAAC GCGAATTAA TCCGGCATAA ACTTAAAGAG GTATTCATC
1051 TGTGCCAAAG CCTCCAGAGG ACAAGCCAGA AGATGTACAT ACAAGTCATC
1101 CATTAAACA AAGAAGAAGA ATATAGAGTG CCAGCAGCAA CTTAGTATTT
1151 TCTAAAAGA ACATTTATTA TTTATTTTA GCCTGTCTAT TTAATCTTTC
1201 AAGAGATTTT ACTGCTGGTA TTTTGTGATG CACTCCTCTT TGAATTTTCA
1251 TTCAAGCCAT TTGTCTAAAG TCATTCTTT GTTTTGGG AGATGGAGTC
1301 TTGCTCTGTT GCCCAGGCTG GAATGCAGTG GCGTGATCTC GGCCTCACTGC
1351 AACCTCCACC TCCCGGGTTC AAGCGATTCT CTTGCCCTCAG CCTCTGAGT
1401 ATCTGGGATT ACAGCGGTGC ACCACCATGC CTGGCTAAGT TTTGTGTTTT
1451 TTTTAGTAGA GATGGGTTT CACCATATTG GTCAGGCTGG TCTCGAACTC
1501 CTGACCTTGT GATACACCTG CCTCAGCCTC CCAAGGGGAT GAGCCACCGC
1551 GCCTGGCCCA TTTCTCTTT TTTTGACCCA TACTTAATGT TGCAGAAACT
1601 ATTCTGTGTA TAACATTATC TCTCATGTAC AGTAATTATA TGTAAATTAA
1651 TTGAAGCAAA TATGGAACT TTACAATAGA AATAAAGATA GGCAGCCAGC
1701 GTCTGTTTCC AATTATAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry AC005156 from database EMBL:
Homo sapiens PAC clone DJ1099C19 from 7q21-q22, complete sequence.
Score = 2897, P = 2.4e-154, identities = 583/586
2 exons covering Bp 465-1723

Medline entries

No Medline entry

Peptide information for frame 3

 ORF from 24 bp to 1103 bp; peptide length: 360
 Category: similarity to unknown protein

```

1 MASSGGEPS LFDHHVQRAV CDTRAKYREG RRPRAVKVYT INLESQYLLI
51 QGVPAVGVMK ELVERFALYG AIEQYNALDE YPAEDFTEVY LIKFMNLQSA
101 RTAKRKMDQ SFFGGLLHVC YAPEFETVEE TRKKLQMRKA YVVKTTENKD
151 HYVTKKKLVT EHKDTEDFRQ DFHSEMSGFC KAALNTSAGN SNPYLPYSCE
201 LPLCYFSSKC MCSSGGPVDR APDSSKDRN HHKTMGHYNH NDSLRLTQIN
251 SLKNSVACPG AQKAITSSA VDRFMPRTTQ LQERKRRRED DRKLGTFLLQT
301 NPTGNEIMIG PLLPDISKVD MHDDSLNTTA NLIRHKLKEV FHLCSLQRT
351 SQKMYIQVIN
  
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_23o5, frame 3

TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence., N = 2, Score = 114, P = 3.6e-11

>TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC F15K20 genomic sequence, complete sequence.
 Length = 227

HSPs:

Score = 114 (17.1 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 21/41 (51%), Positives = 29/41 (70%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPEFETVEETRKKLQMRKAYVV 143
 AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+
 Sbjct: 51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVL 91

Score = 107 (16.1 bits), Expect = 2.6e-10, Sum P(2) = 2.6e-10
 Identities = 50/191 (26%), Positives = 83/191 (43%)

Query: 103 AKRKMDEQSFFGGLLHVCYAPEFETVEETRKKLQMRKAYVVKTENKDHVYTKKKLVTEH 162
 AKRK+DE SF G L + YAPE+E V +T+ KL+ R+ V+ + T + VT+
 Sbjct: 51 AKRKLDESSFLGNRLQISYAPEYENVNDTKDKLESRRKEVLARLNPQKEKSTSQ--VTKL 108

Query: 163 KOTEDFRQDFHSEMSGFCFCAALNTSAGNSNPYLPYSCPLCYFSSKCMCSSGGPVDRAP 222
 + D S + + GN+ P S + YF+S M + V
 Sbjct: 109 AGPALTQTDNVSSQRREMEYQFHR--GNA-PVTRVSSDQE--YFASSSMNQTVKTV---- 159

Query: 223 DSSKDGRRHHKTMGHYNHNDLSLRKTQINSLKNSVACPGAQKAITSSAEDRFMPRTTQLQ 282
 K + + + +H + + + N + P +Q S R P ++Q+Q
 Sbjct: 160 -REKLNKTREENISSLSHCKQIEESG-NQKRLQ---PSSQTQPEESGNQKRLQP-SSQIQ 213

Query: 283 -ERKRRREDDRK 293
 + KR R D+R+
 Sbjct: 214 PDLKRTRVDNRR 225

Score = 102 (15.3 bits), Expect = 3.6e-11, Sum P(2) = 3.6e-11
 Identities = 22/55 (40%), Positives = 38/55 (69%)

Query: 26 KYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMKELVERFALYGAIQY--NALDE 80
 +Y++ P AV+VYT+ ES+Y++++ VPA+G +L+ F YG +E++ LDE
 Sbjct: 3 RYKD-ETP-AVRVYTVCDSESRMIVRNPALGCGDDLMLRFMTYGEVEEFAKRLDE 57

Pedant information for DKFZphfbr2_23o5, frame 3

Report for DKFZphfbr2_23o5.3

```

[LENGTH]      360
[MW]           41105.85
[pI]           8.89
[HOMOL]        TREMBL:AC005824_10 gene: "F15K20.11"; Arabidopsis thaliana chromosome II BAC
F15K20 genomic sequence, complete sequence. 5e-12
[PROSITE]      1
[PROSITE]      MYRISTYL 2
[PROSITE]      CK2_PHOSPHO_SITE 7
  
```

[illegible]

PS00001	185->189	ASN_GLYCOSYLATION	PDCC00001
PS00001	241->245	ASN_GLYCOSYLATION	PDCC00001
PS00001	327->331	ASN_GLYCOSYLATION	PDCC00001
PS00005	99->102	PKC_PHOSPHO_SITE	PDCC00005
PS00005	102->105	PKC_PHOSPHO_SITE	PDCC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDCC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDCC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDCC00005
PS00005	224->227	PKC_PHOSPHO_SITE	PDCC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDCC00005
PS00005	251->254	PKC_PHOSPHO_SITE	PDCC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDCC00005
PS00006	4->8	CK2_PHOSPHO_SITE	PDCC00006
PS00006	10->14	CK2_PHOSPHO_SITE	PDCC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDCC00006
PS00006	224->228	CK2_PHOSPHO_SITE	PDCC00006
PS00006	266->270	CK2_PHOSPHO_SITE	PDCC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDCC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDCC00006
PS00008	5->11	MYRISTYL	PDCC00008
PS00008	260->266	MYRISTYL	PDCC00008
PS00009	29->33	AMIDATION	PDCC00009

194

DKFZphfbr2_2a2

group: brain derived

DKFZphfbr2_2a2.3 encodes a novel 167 amino acid protein with weak similarity to human 52K autoantigen Ro/SS-A

The novel protein contains a C3HC4 Zinc finger "RING finger" motive. This domain is probably involved in mediating protein-protein interactions. Proteins containing a RING-finger are: mammalian V(D)J recombination activating protein (RAG1), mouse rpt-1, human rfp, human 52 Kd Ro/SS-A protein and others.

No informative BLAST results; no predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to 52K autoantigen Ro/SS-A - human

complete cDNA, complete cds, few EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1376 bp

Poly A stretch at pos. 1355, polyadenylation signal at pos. 1340

```
1 GGGGACTCCA AATTAGAAAG GGGACGTCTA GTGGGTGCC CGGGAGGGGT
51 GGGGGAGCG GTCTGGAAG TAATCTGTCC TCTGTCGCC GGAACGGCG
101 AGGTAGTTCC TTCGCGTGG AGAGACCTGG AATGGCCAAA TATCAAGGTG
151 AAGTTCAAAG TTGAAACTG GATGATGATT CAGTTATAGA AGGAGTAAGC
201 GACCAAGTAC TTGTGGCAGT TGTGGTCAGT TTCGCTTTGA TTGCTACCCT
251 GGTATATGCA CTTTTCAGAA ATGTACATCA AAACATTAC CCAGAAAACC
301 AGGAGCTAGT AAGGGTACTT CGAGAACAGC TTCAAACAGA ACAGGATGCA
351 CCTGCTGCCA CTCGACAGCA GTTCTACACT GACATGTACT GTCCCATCTG
401 CCTGCACCAA GCCTCCTTCC CGGTGGAGAC CAACGTGGA CATCTTTTTT
451 GTGGTGCCTG CATTATTGCT TACTGGCGAT ATGGTTCATG GCTTGGGGCA
501 ATCAGTTGTC CAATCTGTAG ACAAACGGTA ACCTTACTCC TAACAGTATT
551 TGGTGAAGAT GATCAGTCTC AGGATGTTCT GAGATTGCAT CAGGATATTA
601 ATGATTATAA CCGGAGATTC TCAGGGCAAC CCTGATCTAT TATGGAGAGA
651 ATTATGGATC TACCCACTTT ACTGAGGCAT GCATTGAGG AAATGTTTTT
701 AGTCGGGGGC CTTTCTGGA TGTTCGCAT CAGGATAATA CTTTGTATA
751 TGGGAGCTTT TTTCTATCTT ATATCACCTC TAGATTTTGT ACCTGAAGCC
801 TTGTTTGGAA TTCTAGGCTT TCTAGATGAT TTCTTTGTCA TCTTTTATT
851 GCTTATCTAC ATCTCTATTA TGTATCGAGA AGTGATAACC CAAAGGCTAA
901 CTAGATGAAA AAGGAAACAA AACTGACTTT ACTAGGATAT CTGAGCTAAT
951 CTAGAACATC AACAGAAGG ACCCATGGCA GTATAAGCA ATGAAGCAAT
1001 GGAGTATTAT CTCACAATA TAAACCACT ATAAGACAAA CATTTGATTA
1051 TCATTTGACA AATACCTAGG TATAACTGGA ATTTTCATGT TGAAGTTCT
1101 AATATTAAAT TTAGAATTAT AATGATCTAC AGTTGTATCT TGATTTCTATG
1151 TTGCTCTGAA AAAATATGGA ATTATATAAA AAGGGATGCT TTTATATATT
1201 TTTCTTTTCC CCAGAATTAC TTAGATTAAT TAGATGTATA GTAAATATTT
1251 GTTAAATGTC AGTTTATCCA TCTTATCCTT CTCAGCAGGT ACCTATATGA
1301 TAATATATAG CTGTGAAACT CATCTAAATA TTTTGTTC ATATAAATAT
1351 TATATACTAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 132 bp to 632 bp; peptide length: 167
Category: similarity to known protein
Classification: unset

Prosite motifs: ZINC_FINGER_C3HC4 (102-112)

```

1 MAKYQGEVQS LKLDSDSVIE GVSDQVLVAV VVSFALIATL VYALFRNVHQ
51 NIHPENQELV RVLREQLQTE QDAPAAATROQ FYTDMYCPIC LHQASFPVET
101 NCGHLFCGAC IAYWRYGSW LGAISCPICR QTVTLTLLTVF GEDDQSQDVL
151 RLHQDINDYN RRFSGQP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2a2, frame 3

TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A, N = 1, Score = 194, P = 2e-15

PIR:T05222 hypothetical protein F17I5.130 - Arabidopsis thaliana, N = 1, Score = 159, P = 1.4e-10

TREMBLNEW:AB025011.1 gene: "TRIF"; product: "Trif-d"; Mus musculus mRNA for Trif-d, cOmplete cds., N = 1, Score = 108, P = 2.6e-06

PIR:A37241 52K autoantigen Ro/SS-A - human, N = 1, Score = 115, P = 5e-05

>TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A Length = 283

HSPs:

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 52/149 (34%), Positives = 78/149 (52%)

```

Query:   16 DSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELVRLREQLQTEQDAPA 75
          D +E ++ Q+ +A+ V F ++ + A Q E R Q+ T++
Sbjct:   41 DPDVE-LATQITMAIAVIF-IVKAIFDAWQSRRRQRAASRMENAE--RNQIITQRRISE 96
          A Q + CPICL ASFPV T+CGH+FC CII YW+ + C +CR T
Query:   76 ATRQQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSWLGA-ISCPCICRQTVT 134
          A Q + CPICL ASFPV T+CGH+FC CII YW+ + C +CR T
Sbjct:   97 ALHQSSHE---CPICLANASFPVLTDCGHIFCCECI IQYQQSKAIVTPCDAMCRSTFY 153
          +LL V G +++ D ++ + I+DYNRRFS
Query:   135 LLLTV---FGEDDQSQDVLRLHQ-DINDYNRRFS 164
          +LL V G +++ D ++ + I+DYNRRFS
Sbjct:   154 MLLPVHWPTMGTSEETDDHIQENNIRIDYNNRRFS 188

```

Pedant information for DKFZphfbr2_2a2, frame 3

Report for DKFZphfbr2_2a2.3

```

[LENGTH]      167
[MW]           18941.65
[pI]           4.91
[HOMOL]        TREMBL:CEY38F1A_8 gene: "Y38F1A.2"; Caenorhabditis elegans cosmid Y38F1A 1e-13
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       30.19 peroxisomal organization [S. cerevisiae, YDR265w] 1e-04
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR323c] 2e-04
[BLOCKS]       BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE]      ZINC_FINGER_C3HC4 1
[PFAM]         zinc finger, C3HC4 type (RING finger)
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY 6.59 %

```

```

SEQ  MAKYQGEVQSLKLDSDSVIEGVSDQVLVAVVVSFALIATLVYALFRNVHQNIHPENQELV
SEG  .....XXXXXXXXXXXXX.....
lcmd-.....

SEQ  RVLREQLQTEQDAPAAATROQFYTDMYCPICLHQASFPVETNCGHLFCGACIIAYWRYGSW
SEG  .....HHHHHHBTTTTEETTTEETTEEEHHHHH---HHHHH
lcmd-.....

SEQ  LGAISCPICRQTVTLTLLTVFGEEDQSQDVLRLHQDINDYNRRFSGQP

```

WO 01/12659

PCT/IB00/01496

SEG
lrmd- HCCB-TTTT.....

Prosites for DKF2phfbr2_2a2.3

PS00518 102->112 ZINC_FINGER_C3HC4 PDOC00449

Pfam for DKF2phfbr2_2a2.3

HMM_NAME Zinc finger, C3HC4 type (RING finger)
HMM *CPICFCTFQldyPWPFFdePmMlPCgHsFCypCIrrW.....CP
CPIC L+ P++++CGH+FC +CI+ + CP
Query 87 CPIC-----LHQ---ASFPVETNCGHLFCGACIIAYWRYGSLGAISCP 127
HMM mC*
+C
Query 128 IC 129

DKFZphfbr2_2b17

group: transmembrane protein

DKFZphfbr2_2b17 encodes a novel 285 amino acid protein with similarity to D. melanogaster 30K protein.

The protein contains 3 transmembrane regions.
No informative BLAST results; no predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to Drosophila hypothetical 30K protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 3

Sequenced by Qiagen

Locus: unknown

Insert length: 1426 bp

Poly A stretch at pos. 1345, polyadenylation signal at pos. 1330

```
1 GGGGGTATTT CCAAGGACTC CAAAGCGAGG CCGGGGACTG AAGGTGTGGG
51 TGTCGAGCCC TCTGGCAGAG GGTAAACCTG GGTCAAATGC ACGGATTCTC
101 ACCTCGTACA GTTACGCTCT CCCGCGGCAC GTCCGCGAGG ACTTGAAGTC
151 CTGAGCGCTC AAGTTGTGCC GTAGGTCGAG AGAAGGCCAT GGAGGTGCCG
201 CCACCGGCAC CGCGGAGCTT TCTCTGTAGA GCATTGTGCC TATTCCCCG
251 AGTCTTTGCT GCCGAAGCTG TGAAGTCCGA TTCGGAAGTC CTTGAGGAGC
301 GTCAGAAAGCG GCTTCCCTAC GTCCGAGAGC CCTATTACCC GGAATCTGGA
351 TGGGACCGCC TCCGGGAGCT GTTTGGCAAA GATGAACAGC AGAGAATTC
401 AAAGGACCTT GCTAATATCT GTAAGACGGC GCCTACAGCA GGCATCATTG
451 GCTGGGTGTA TGGGGGAATA CCAGCTTTTA TTCATGCTAA ACAACAATAC
501 ATTGAGCAGA GCCAGGCAGA AATTTATCAT AACCGGTTTG ATGCTGTGCA
551 ATCTGCACAT CGTGTGCCA CACGAGGCTT CATTCTGTAT GGCTGGCGCT
601 GGGGTTGGAG AACTGCAGTG TTTGTGACTA TATTCAACAC AGTGAACACT
651 AGTCTGAATG TATACCGAAA TAAAGATGCC TTAAGCCATT TTGTAATTGC
701 AGGAGCTGTC ACGGGAAGTC TTTTtaggat AAACGTAGGC CTGCGTGGCC
751 TGGTGGCTGG TGGCATAATT GGAGCCTTGC TGGGCACTCC TGTaggagGC
801 CTGCTGATGG CATTTCAGAA GTACTCTGGT GAGACTGTTT AGGAAAGAAA
851 ACAGAAAGGAT CGAAAGGCAC TCCATGAGCT AAAACTGGAA GAGTGGAAAG
901 GCAGACTACA AGTTACTGAG CACCTCCCTG AGAAATGAAG AAGTAGTTTA
951 CAGSAAGATG AACCTGAGAA TGATGCTAAG AAAATTGAAG CACTGCTAAA
1001 CCTTCCTAGA AACCTTCAG TAATAGATAA ACAAGACAAG GACTGAAAGT
1051 GCTCTGAAGT TGAAGCTCAC TGGAGAGCTG AAGGAGCTG CCATGTCCGA
1101 TGAATGCCAA CAGACAGGCC ACTCTTTGGT CAGCCTGCTG ACAAAATTAA
1151 GTGCTGCTAC CTCTGGTGGC AGTGGCTTGC TCTGTCTCTT TTCCTTTCTT
1201 TTTAACTAAG AATGGGGCTG TTGTAAGTCT ACTTTACTTA TCCTTAAATT
1251 TAAATACATA CTATGTTTG TATTAATCTA TCAATATATG CATACATGAA
1301 TATATCCACC CACCTAGATT TTAAGCAGTA AATAAACAT TTCGCAAAAG
1351 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HSG19630 from database EMBL:
human STS A001T27.
Score = 961, P = 1.2e-36, identities = 193/194

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 189 bp to 1043 bp; peptide length: 285
Category: similarity to unknown protein

```

1 MEVPPPPAPRS FLCKALCLFP RVFAAAEAVTA DSEVLIEERQK RLPYVPEFY
51 PESGWDLRLK LFGKDEQQRRI SKDLANICKT AATAGYIGRW YGGPFAVFIH
101 KQYIEQSQA LYYHNRFDAP SQASHRAACT FTIRYGWRWG TAPFTVFTIN
151 TWNTSLNVYR NKDALSHFVI AGAVTGSFLR INVGLRGLVA GGIIGALLGT
201 PVGGLLMAFY KYSGETVQER KQKDRKALHE LKLEWKGRL QVTEHLPKI
251 ESSLQDEQDE NDANKIEALL NLPNPSVID QKQD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2b17, frame 3

PIR:JQ1024 hypothetical 30K protein (DmRPL40 5' region) - fruit fly (*Drosophila melanogaster*), N = 1, Score = 312, P = 6.1e-28

```
>PIR:JQ1024 hypothetical 30K protein (DmRPL40 5' region) - fruit fly
(Drosophila melanogaster)
Length = 261
```

HSPs :

Score = 312 (46.8 bits), Expect = 6.1e-28, P = 6.1e-28
Identities = 68/231 (29%), Positives = 125/231 (54%)

Query: 30 ADSEVLEERQKRLPYVPEPYYPESGWDRRLRELFKGDEQQRIKDLANICKTAATAGIIGW 89
AD V +E + ++ E+G +RL+++F DE I +L ++ + +IG
Sbjct: 23 ADEIVDKENKTYKAFLASKPPEETGLERLQMFITIDFGSIFSELNSVYQAGFLGFLIGA 82

Query: 90 VYGGIPAFIHAQQYIEQSQAIEYHNRFDAVQSAHRAATRGFIYGRWGWRTAVFVTIF 149
+YGG+ A ++E +QA + + FDA + T F + G++WGWR +F T +
Sbjct: 83 IYGGVTSRVAYMNFMENNQATAFKSHFDAKKKLQDQFTVNFAKGGFKWGWVRVGLFTTSY 142

Query: 150 NTVNTSLNVYRNKDALSHFVIAGAVTGSFLFRINVGRLGVAGGIIGALLGTPVGGLLMAF 209
+ T ++VYR K ++ . ++ AG++TGSF++++GLRG+ AGGIIG LG G +
Sbjct: 143 FGIITCSVYRGKSSIIYEYLAAGSITGSFLYKSLGLRGMAGGIIGGFLGGVAGVTSLL 202

```
Query: 210 QKYSGETVQERKQKDRKALHELKLEEWKGRQLQVTEHLPEKIESSLQEDEPE 260
      K SG +++E ++ ++K RL E++ + + +++ PE
Sbjct: 203 MKASGTSMEE-----VRYWQYKWLDRDENIQQAFKKLTEDENPE 242
```

Pedant information for DKF2phfbr2_2b17, frame 3

Report for DKFZphfbr2_2b17.3

```

[LENGTH]      285
[MW]           32177.88
[pI]           8.65
[HMOL]        PIR:JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila
melanogaster) 7e-20
[PROSITE]     MYRISTYL      7
[PROSITE]     CK2_PHOSPHO_SITE      5
[PROSITE]     ASN_GLYCOSYLATION      1
[KW]          SIGNAL_PEPTIDE 25
[KW]          TRANSMEMBRANE  3
[KW]          LOW_COMPLEXITY      5.96 %

```

```

SEQ      MEVPPPPAPRSFLRCALCLFPRVFAAEAVTADSEVLEERQKRLPYVPEPYYPESGDWRDLRE
SEG
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      LFGKDEQQRISKDLANICKTAATAGIIGVWYGGIPAFIHAQQYIEQSQAEIYHNRFDV
SEG
PRD      hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      QSAHRAATRGFIYRGWRGWRGRTAVFVTIFNTVNTSLNVYRNKDALSHFVIAGAVTGSFLR
SEG
PRD      hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....M

SEQ      INVGLRGLVAGGIIGALLGTPVGGLLMAFQKSYGETVQERKQKDRKALHELKLEEWKGRGL

```

```

SEG      ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
MEM      mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

```

```

SEQ      QVTEHLPEKIESSLQEDEPENDAKKIEALLNLPNPSVIDKQDKD
SEG      .....
PRD      cccccccchhhhhccccccchhhhhhhhhhhccccceeeccccc
MEM      .....

```

Prosites for DKFZphfbr2_2b17.3

PS00001	153->157	ASN_GLYCOSYLATION	PDOC00001
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	108->112	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2b17.3)

DKF2phfbr2_2b5

group: cell structure and motility

DKF2phfbr2_2b5 encodes a novel 957 amino acid protein with strong similarity to collagens.

The novel protein contains the typical (xxG)n repeat of collagen proteins and a Pfam von Willebrand factor type A domain. Therefore, the protein seems to be a new collagen alpha chain.

The new protein can find application in modulation of connective tissue, bone and cartilage development and maintenance.

similarity to collagen proteins

shows typical (xxG)n repeat of collagen proteins
[PFAM] von Willebrand factor type A domain

Sequenced by Qiagen

Locus: /map="6"

Insert length: 4160 bp

Poly A stretch at pos. 4141, polyadenylation signal at pos. 4119

```
1 GGGGGCCCCG TGCAGGGAGA ACGGACTCCG GCGGAGGGC AGCCAATCCG
51 TTTTCAGCGCA GGTCTTGCTC GGGTTGGGCT TGCCACTGCC TGGAAACATAC
101 CTGTCCCCCT GCGCAACAC TCAGCTGGCT GCGACCGCAA CCCCAGCCCT
151 GGACACTGCG CCAGGAATCC TAAAACCAA ATATTAGAAC GAAAACAGAA
201 ACATGGCTCA CTATATTACA TTTCTCTGCA TGGTTTGGT GCTGCTCTT
251 CAGAATTCTG TGTAGCTGA AGATGGGAA GTAAGATCAA GTTGTCTGAC
301 TGCTCCGACA GATTTAGTTT TCATCTTAGA TGGCTCTTAT AGTGTGGCC
351 CAGAAAACTT TGAAATAGTG AAAAAGTGGC TTGTCAATAT CACAAAAAAC
401 TTTGACATAG GGCCGAAGTT TATTCAAGTT GGAGTGGTTC AATATAGTGA
451 CTACCCTGTG CTGGAGATTC CTCTCGGAAG CTATGATTCA GGAGAACATT
501 TGACGGCAGC AGTGGAAATCC ATACTCTACT TAGGAGGAAA CACAAAGACA
551 GGGGAAGGCCA TCCAGTTTGC GCTCGATTAC CTTTTTGACA AGTCCTCAGC
601 ATTTCTGACT AAGATAGCAG TGGTACTTAC GGATGGCAAG TCCCAAGATG
651 AGCTCAAGGA TGCAGCTCAA GCAGCAAGAG ATAGTAAGAT AACATTATTT
701 GCTATTGGTG TTGGTTCAGA AACAGAAGAT GCCGAACCTA GAGCTATTGC
751 CAACAAGCCT TCGTCTACTT ATGTGTTTAA TGTGGAAGAC TATATTGCAA
801 TATCCAAAT AAGGGAAGTG ATGAAGCAGA AACTTTGTGA AGAATCTGTC
851 TGTCCAACAC GAATTCAGT GGCAGCTCGT GATGAAAGGG GATTTGATAT
901 TCTTTTGGGT TTAGATGTAA ATAAAAAGGT TAAGAAAAGA ATACAGCTTT
951 CACCAAAAAA GATAAAAGGA TATGAAGTAA CATCAAAAGT TGATTATCA
1001 GAACTCACA GCAATGTTTT CCCAGAAGGT CTTCCTCCAT CATATGTATT
1051 TGTGTCTACT CAAAGATTAA AAGTCAAGAA AATTTGGGAT TTATGGAGAA
1101 TATTAACTAT TGATGGAAGG CCACAAATAG CAGTTACCTT AAATGGTGTG
1151 GACAAAATCT TATTATTTAC AACAAACCAGC GTAATTAATG GCTCACAAGT
1201 GGTTACCTTT GCTAACCTC AAGTTAAGAC GTTGTTTGAT GAAGGCTGGC
1251 ACCAAATTCG TCTCTTAGTA ACAGAACAAG ATGTGACTTT GTATATTGAT
1301 GACCAACAAA TTGAAAACAA GCCCTTACAT CCAGTTTTAG GGATCTTGAT
1351 CAATGGGCAA ACCCAAATTG GAAATATTC TGGAAAAGAA GAAACTGTTC
1401 AGTTTGTATG CCAAAAGTTG CGAATCTACT GTGACCCAGA ACAGAACAAC
1451 CGGGAGACAG CATGTGAGAT TCCTGGATT AATGGAGAGT GCCTTAATGG
1501 TCCAGTGAT GTAGGTTCAA CTCCAGCTCC CTGTATTGT CCTCCGGGAA
1551 AACCAGGACT TCAAGGCCCC AAAGGTGACC CTGGACTGCC TGGGAACCCCT
1601 GGTACCTCTG GACAACTGG TCAAGATGGT AAGCTGGAT ATCAGGGAAT
1651 TGCAGGGACA CCAGGTGTTT CAGGATCTCC AGGAATACAA GGAGCTCGAG
1701 GACTACCAGG TTACAAAAGGA GAACCAAGGC GAGATGGTGA CAAGGGTGAT
1751 CGTGGACTTC CTGTTTTTCC TGGGCTTCAT GGCATGCCAG GATCAAAGGG
1801 TGAATGGGT GCCAAAGGAG ACAAAGGATC ACCTGGATT TATGGCAAAA
1851 AGGGTGCAAA AGGTGAAAAG GGAATGCTG GCTTCCCTGG CCTCCCTGGA
1901 CCTGCTGGAG AACCAGGAAG ACATGGAAG GATGGATTAA TGGGTAGTCC
1951 CGGTTTCAAG GGAGAAGCAG GATCCCTGG TGCTCCGGGG CAGGATGGAA
2001 CACGGGGAGA GCCTGGAATC CCAGGATTTT CTGGAACCG AGGATTAATG
2051 GGCCAAAAGG GAGAAATTGG GCCTCCAGGA CAGCAAGGAA AAAAAGGAGC
2101 CCCAGGGATG CCTGGTTTAA TGGGAAGCAA TGGCTCACCA GGCCAGCCTG
2151 GAACACCGGG ATCTAAGGGA AGCAAAGGTG AACCTGGAAT TCAAGGGATG
2201 CCTGGGGCTT CAGGGCTCAA GGGAGAACCA GGAGCAACGG GTTCCCCAGG
2251 AGAACCAGGA TACATGGGTT TACCCGGGAT TCAAGGAAAA AAGGGGGACA
2301 AAGGAAATCA AGGTGAAAAA GGTATTCAGG GTCAAAAGGG AGAAAATGGA
2351 AGACAGGGAA TTCCAGGGCA ACAGGGAATT CAAGGCCATC ATGGTGCAAA
2401 AGGAGAGAGA GGTGAAAAGG GAGAACCTGG TGTCGAGGT GCCATTGGAT
2451 CAAAAGGAGA ATCTGGGGTG GATGGCTTGA TGGGGCCCCG AGGTCTTAAG
2501 GGGCAACCTG GGGATCCAGG TCCTCAGGGA CCCCAGGTT TGGATGGGAA
2551 GCCCGGAAGA GAGTTTTCAG AACAAATTAT TCGACAAAGT TGCACAGATG
```

```

2601 TAATAAGAGC CCAGCTACCA GTCTTACTTC AGAGTGAAG AATTAGAAAT
2651 TGTGATCATT GCCTGTCCCA ACATGGCTCC CCGGGTATTC CTGGGCCACC
2701 TGGTCCGATA GGCCAGAGG GTCCAGAGG ATTACCTGGT TTGCCAGGAA
2751 GAGATGGTGT TCCTGGATTA GTGGGTGTCC CTGGACGTCC AGGTGTCAGA
2801 GGATTAAAG GCCTACCAGG AAGAAATGGG GAAAAAGGGA GCCAAGGGTT
2851 TGGGTATCCT GGAGAACAAG GTCCTCCTGG TCCCCAGGT CCAGAGGGCC
2901 CTCCTGGAAT AAGCAAAGAA GGTCTCTCAG GAGACCCAGG TCTCCCTGGC
2951 AAAGATGGAG ACCATGGAAA ACCTGGAATC CAAGGGCAAC CAGGCCCCCC
3001 AGGCATCTGC GACCCATCAC TATGTTTTAG TGTAAATGCC AGAAGAGATC
3051 CGTTCAGAAA AGGACCAAAC TATTAGTGTC TGATGCCTCA TTCAGCAGCC
3101 TAGGCATGGT GCTTTTTCTG TGGTCTTTTG CATCTCAGGA AGATAACCAA
3151 CAGTATCCCT TGAAAAGAAA CTTAAGTACC TCGGTGTTTT TATTTTTTTT
3201 TTCTTATGGA AAAAAATATA AAAGATCACA TATACTGATT TTAAGGGCTC
3251 CTCAGTCATT TGGAGCCCTT GGATTAGCAG CATTAAATTA ATCTCAAGGG
3301 TTTCTTGTA AGTCCATTTA TGTTAATCAA AGTTGAATAT AAAAAATCCAC
3351 CATTGCCTGT TAGCCAGTCA GTTTTAGTCA CTGTGAAATA TTTCACATTC
3401 AGCCTCCATG CAGTAGAGAT TTGAGTTTAA TTTCATGTCC ATGTGACTTT
3451 CATGTTTCTC ATCTCATAGC TCATGCTACT ACATAAGCCA AAACATGTAT
3501 CTCATCATTTG GAAGTAAGAT CAGGGCTGAT ATTCACCTGG GATAGACAGT
3551 ATTTGGTGAAC TACTCATTTA CTACAGTGTC TCAGCCTTGA TAAAGGGCAG
3601 TGGATTGGCT GTTGTTCGGT GTTGTGAATA GCACCTCTGA ATAAGATTAG
3651 AGTGTTCCTT AATTCATTTC AAACCTCTAA ATTAGATTAA TGGTGGTGCT
3701 AAGAAAGAGT ATTAATTACT TTGGGAATGG TCAAAATTAA CATTAAAAAC
3751 ATTTTAGACA AAAAGTTTCA TTGTACATTC AAAGAAAATG TAAGTTTGGA
3801 AGTACTAAAA GACTATTTTA TACTTGTTGA TTAATCGGAA TGTTTGTGTG
3851 ATGCCTTCAT TTTCCATTTC ACTTATATGT GCATGTCCAT ATATGTTAAT
3901 TTTTATTGTA GCAAAGCTAA TGGAAATAAA GCTAATGCTC TAGTTGAAAG
3951 AAAAGGAAAA CTCCTGAAAT CCTAGAATGT CTTGTTATTT TTAGCTGACT
4001 GTAAATATT ATGAACAGTC TTTGTGTATT GTGCTTAATG CTTTGTGAAG
4051 AAACAGAAAT TGAAATATTT CATCCTTGTC ATGCTCAAAA TTTTGTTACA
4101 TGCTTGTAT TCAGAGTATA ATAAAGTTT GTACAGGCCT GAAAAAATAA
4151 AAAAAAATAA

```

BLAST Results

Entry HS682J15 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 682J15
 Score = 6240, P = 0.0e+00, identities = 1256/1263
 13 exons matching Bp 2015-4118

Entry HS708F5 from database EMBLNEW:
 Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 708F5
 Score = 2775, P = 1.0e-221, identities = 739/912
 10 exons matching Bp 5-1745

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 203 bp to 3073 bp; peptide length: 957
 Category: similarity to known protein

```

1 MAHYITFLCM VLVLLQNSV LAEDGEVRSS CRTAPTDLVF ILDGYSYSVGP
51 ENFEIVKWL VNITKNFDIG PKFIQVGVVQ YSDYPVLEIP LGSYDSGEHL
101 TAAVESILYL GGNTKTGKAI QFALDYLFDK SSRFLTKIAV VLTGKSQDD
151 VKDAAQAARD SKITLFAIGV GSETEDAELR AIANKPSSSTY VFYVEDYIAI
201 SKIREVMKQK LCEESVCPTR IPVAARDEGR FDILLGLDVN KVKVKRIQLS
251 PKKIKGYEVT SKVDLSELTS NVFPEGLPPS YVFVSTQRFK VKKIWDLWRI
301 LTIDGRPQIA VTLNGVDKIL LFTTTSVING SQVVTFANPQ VKTLFDEGWH
351 QIRLLVTEQD VTLYIDDOQI ENKPLHPVLG ILINGQTQIG KYSGKEETVQ
401 FDVQKLRIYC DPEQNNRETA CEIPGFNGEC LNGPSDVGST PAPCICPPGK
451 PGLQGPKGDP GLPGNPGYPG QPGQDGKPGY QGIAGTPGVP GSPGIQGARG
501 LPPGYKGEPR DGDKGDRGLP GFPLHGMPLG SKGEMGAKD KGSPPGYGKK
551 GAKGEKGNAG FPGLPGPAGE PGRHKDGLM GSPGFKGEAG SPGAPQDGT
601 RGEPIPGFPP GNRGLMGQKG EIGPPGQQGK KGAPGMPGLM GSNGPSGQPG
651 TPGSKGSKEG PGIQGMPGAS GLKGEPGATG SPGEPIGYML PGIQGKKGDK
701 GNQGEKGIQG QKGENGROGI PQQQGIQGHG GAKGERGEKG EPGVRGAIGS
751 KGESGVDLGM PGAPGKGQPG DPGPQGPPGL DGKPGREFSE QFIRQVCTDV
801 IRAQLPVLLQ SGRIRNCDHC LSQHGSPGIP GPPGPIGPEG PRGLPGLPGR

```

851 DGVFGLVGVF GRPGVRGLKG LPGRNGEKGS QGFGYPGEQG PPGPPGPEGP
 901 PGISKEGPPG DPGLPGKGDG HGKPGIQGQP GPPGICDPSL CFSVIARRDP
 951 FRKGPNY

BLASTP hits

Entry HSCOL7A1x_1 from database TREMBL:
 gene: "COL7A1"; product: "collagen type VII"; Homo sapiens (clones:
 CW52-2, CW27-6, CW15-2, CW26-5, 11-67) collagen type VII intergenic
 region and (COL7A1) gene, complete cds.
 Score = 949, P = 3.4e-122, identities = 237/553, positives = 281/553

Entry CA17_HUMAN from database SWISSPROT:
 COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC
 COLLAGEN). >TREMBL:HSCOL7A1_1 gene: "COL7A1"; product: "alpha-1 type
 VII collagen"; Human alpha-1 type VII collagen (COL7A1) mRNA, complete
 cds.
 Score = 949, P = 3.6e-122, identities = 237/553, positives = 281/553

Alert BLASTP hits for DKFZphfbr2_2b5, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2b5, frame 2

Report for DKFZphfbr2_2b5.2

[LENGTH] 957
 [MW] 99413.38
 [pI] 8.49
 [HOMOL] PIR:A40020 collagen alpha 1(XII) chain precursor - chicken 9e-90
 [BLOCKS] BL01119B Copper-fist domain proteins
 [BLOCKS] BL00313B
 [BLOCKS] BL01113A Clq domain proteins
 [BLOCKS] BL00420A Speract receptor repeat proteins domain proteins
 [SCOP] dlzoob_ 3.45.1.1.1 Integrin CD11a/CD18 (LFA-1) [Human (Hom 2e-58
 [SCOP] dlido_ 3.45.1.1.2 Integrin CR3 (CD11b/CD18), alpha subunit [Huma 8e-62
 [EC] 3.1.1.7 Acetylcholinesterase 7e-24
 [PIRKW] blocked amino end 1e-43
 [PIRKW] duplication 7e-46
 [PIRKW] cornea 1e-35
 [PIRKW] lung 2e-40
 [PIRKW] leukocyte 1e-42
 [PIRKW] skin 1e-40
 [PIRKW] transmembrane protein 1e-37
 [PIRKW] cartilage 3e-59
 [PIRKW] hydroxylysine 4e-62
 [PIRKW] connective tissue 3e-43
 [PIRKW] triple helix 5e-82
 [PIRKW] homotrimer 2e-37
 [PIRKW] bone 6e-40
 [PIRKW] Alport syndrome 1e-42
 [PIRKW] laminin binding 2e-40
 [PIRKW] liver 2e-40
 [PIRKW] glycoprotein 5e-82
 [PIRKW] carboxylic ester hydrolase 7e-24
 [PIRKW] disulfide bond 7e-46
 [PIRKW] cell binding 7e-46
 [PIRKW] heterotrimer 4e-62
 [PIRKW] calcium binding 8e-28
 [PIRKW] alternative splicing 5e-82
 [PIRKW] coiled coil 5e-82
 [PIRKW] basement membrane 7e-46
 [PIRKW] trimer 5e-82
 [PIRKW] pyroglutamic acid 3e-43
 [PIRKW] hydroxyproline 4e-62
 [PIRKW] extracellular matrix 5e-82
 [PIRKW] chondroitin sulfate proteoglycan 6e-41
 [PIRKW] sulfoprotein 7e-39
 [PIRKW] kidney 1e-42
 [PIRKW] angiogenesis inhibitor 6e-36
 [PIRKW] Ehlers-Danlos syndrome 2e-40
 [SUPFAM] fibronectin type III repeat homology 5e-82
 [SUPFAM] scavenger receptor cysteine-rich domain homology 1e-37
 [SUPFAM] C-type lectin homology 6e-30
 [SUPFAM] collagen alpha 2(I) chain 5e-40
 [SUPFAM] collagen alpha 1(I) chain 6e-44


```

[SUPFAM] fibrillar collagen carboxyl-terminal homology 6e-44
[SUPFAM] animal Kunitz-type proteinase inhibitor homology 2e-38
[SUPFAM] fibronectin type II repeat homology 6e-21
[SUPFAM] complement C1q carboxyl-terminal homology 1e-38
[SUPFAM] collagen alpha 3(VI) chain 2e-31
[SUPFAM] collagen alpha 1(IV) chain 7e-46
[SUPFAM] collagen alpha 1(VI) chain 2e-37
[SUPFAM] von Willebrand factor type C repeat homology 6e-44
[SUPFAM] unassigned collagens 4e-62
[SUPFAM] von Willebrand factor type A repeat homology 5e-82
[SUPFAM] collagen alpha 1(XIV) chain 5e-82
[SUPFAM] pulmonary surfactant protein D 6e-30
[SUPFAM] collagen alpha 1(V) chain 7e-39
[SUPFAM] collagen alpha 1(VIII) chain 1e-38
[SUPFAM] EGF homology 1e-35
[PROSITE] AMIDATION 3
[PROSITE] MYRISTYL 14
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] von Willebrand factor type A domain
[KW] Irregular
[KW] 3D
[KW] SIGNAL_PEPTIDE 23
[KW] LOW_COMPLEXITY 24.24 %

```

SEQ	MAHYITFLCMVLFVLLQNSVLAEDEGEVRSSCRTAPDTDLVFDLGSYSVGPENFEIVKKWL
SEG
latzBCCCEEEEEEECCCCCHHHHHHHHHHH
SEQ	VNITKNFDIGPKFIQVGVVQSDYPVLEIPLGSYDSGEHLTAAVESILYLGNGTKTGKAI
SEG
latzB	HHHHHHCCBTTTTEEEEEEEETTTEEEETTNTTTTHHHHHHHHHHCCCCCCCCCHHHHH
SEQ	QFALDYLFDKSSRFLTKIAVVLTDGKSQDDVKDAAQAARDSKITLFAIGVGSETEDAELR
SEG
latzB	HHHHHHHHCCTTTTTEEEEEEEECCTTTTHHHHHHHHHHHHCEEEEEEECCCCCHHHHH
SEQ	AIANKPSSTYVFYVEDYIAISKIREVMKQKLCEESVCPTRIPVAARDERGFIDILGLDVN
SEG
latzB	HHHGGGGGGGCECHHHHHHHHHHCHHHHHHHH.....
SEQ	KKVKKRIQLSPKKIKGYEVTSKVDLSELTSNVFEPLGPSYVFVSTQRFKVKKIIDLWRI
SEG
latzB
SEQ	LTIDGRPQIAVTLNGVDKILLFTTTSVINGSQVVTFANPQVKTLFDEGWHQIRLLVTEQD
SEG
latzB
SEQ	VTLYIDDQQIENKPLHPVLGILINGQTQIGKYSKEETVQFDVQKLRIYCDPEQNNRETA
SEG
latzB
SEQ	CEIPGFNGECLNGPSDVGSTPAPCICPPGKPLQGPKGDPLPGNPGYPGQPGQDGKPGY
SEGxx
latzB
SEQ	QGIAGTPGVPGSPGIQGARGLPGYKGEPRGDGDKDRGLPGFPGLHGMPSGKMGEMAKGD
SEG	xx.....
latzB
SEQ	KGSPGFYGGKGAKEKGNAGFPGLPGPAGEPRGHGDKGLMGSPGFKEAGSPGAPGQDGT
SEGxxxxxxxxxxxxxxxx.....
latzB
SEQ	RGEPGIPGFPNGRLMGQKGEIGPPGQGGKGPAGMPGLMGSNGSPGQPGTPGSKGSKGE
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
latzB
SEQ	PGIQGMPGASGLKGEPATGSPGEPGYMGLPGIQGKKDGKNQGEKGIQGGKGENGRQGI
SEGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
latzB
SEQ	PGQQGIQGHGAKGERGEKGEPGVRGAIGSKGESVDGLMGPAGPKGQPGDPGPGPPGL
SEG	xxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
latzB
SEQ	DGKPGREFSEQFIRQVCTDVIRAQLPVLLQSGRIRNCDHCLSHQGSPIGPMPGPIGPEG
SEG	xxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

```

latzB .....
SEQ      PRGLPGLPGRDGVPLVGVPGVRGLKGLPGRNGEKGSGQFGYPGEQGPPEGPPGPEGP
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
latzB .....

SEQ      PGISKEGPPGDPGLPGKGDHKGPGIQGPQPGICDPSLCFSVIARRDPFRKGPNY
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
latzB .....

```

Prosites for DKF2phfbr2_2b5.2

PS00001	62->66	ASN_GLYCOSYLATION	PDOC00001
PS00001	329->333	ASN_GLYCOSYLATION	PDOC00001
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	131->134	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	286->289	PKC_PHOSPHO_SITE	PDOC00005
PS00005	393->396	PKC_PHOSPHO_SITE	PDOC00005
PS00005	811->814	PKC_PHOSPHO_SITE	PDOC00005
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	261->265	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	357->361	CK2_PHOSPHO_SITE	PDOC00006
PS00006	393->397	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00006	531->535	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	657->661	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	750->754	CK2_PHOSPHO_SITE	PDOC00006
PS00006	754->758	CK2_PHOSPHO_SITE	PDOC00006
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	112->118	MYRISTYL	PDOC00008
PS00008	236->242	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	527->533	MYRISTYL	PDOC00008
PS00008	596->602	MYRISTYL	PDOC00008
PS00008	638->644	MYRISTYL	PDOC00008
PS00008	650->656	MYRISTYL	PDOC00008
PS00008	653->659	MYRISTYL	PDOC00008
PS00008	665->671	MYRISTYL	PDOC00008
PS00008	743->749	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00009	547->551	AMIDATION	PDOC00009
PS00009	628->632	AMIDATION	PDOC00009
PS00009	694->698	AMIDATION	PDOC00009

Pfam for DKF2phfbr2_2b5.2

HMM_NAME	von Willebrand factor type A domain		
HMM	*DIVFLIDGSdSIGpqNfNrmKDFierMMERMDigPDwIRVGVVQYSdNP		
Query	37	DLVFILDGSYSVGPENFEIVKKWLVNITKNFDIGPKFIQGVVQYSDYP	85
HMM	RqEmrFmFNDYQNKeEILQaIqqMMYwMgggTNTGeAIQYVvrNMFwee		
Query	86	VLE--IPLGSYDSGEHLTAIVESIL-YLGGNTKTGKAIQFALDYLFDKSS	132
HMM	GmRWenvPQVMIIITDGRSQDDIRDpIneMrrmaGIqvFaIGIGNhDNnn		
Query	133	RF----LTKIAVVLTDGKSQDDVKDAAQAARD-SKITLFAIGVGSETE--	175
HMM	WeELReIASePdEdHVfYvDfFeeLdnMqeql*		
Query	176	DAELRAIANKPSSTYVFYVEDYIAISKIREVM	207

DKFZphfbr2_2c1

group: brain derived

DKFZphfbr2_2c1 encodes a novel 697 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 3973 bp

Poly A stretch at pos. 3914, polyadenylation signal at pos. 3900

```
1 GGGGGGATT CGGCGGCGGA AACATGGCGG TCGCGGCGGG GCCGGTAACG
51 GAGAAAGTTT ACGCCGACAC TGGCCTGTAT TAGCGCGTAT GGCCTCGGGC
101 CCTCGTTCCC CAAGGCGTGC CGCCTCCCTG TTCTCAGTCG CAGGCTGAAG
151 CCTGTCTGCT TCTCCTCCTT TTTGGTTTGG TTTTGAACT GACTCCGAGG
201 GTTGGGAGAG CGCGTTGGTG GCGACGGCCG AGTCAGATCA CTATAACAA
251 AATTTCCACA AGAGAAAATG TTGAAATAGG AGTTGCGGAT ACATTGGATA
301 TACTGGATGA AATACAAGCG GTTAATTTT GTAACGTGAG GGAAAAGCCC
351 ACATTGCTGG TTACATGTGT AAATCACTGC GTTATTGCTT TAGTCATTGT
401 CTCTATTTAG CAATGACAAG ACTGGAAGAA GTAAATAGAG AAGTGAACAT
451 GCATTCTTCA GTGCGGTATC TTGGCTATTT AGCCAGAATC AATTTATTGG
501 TTGCTATATG CTTAGGCTCA TACGTAAGAT GGGAAAAAC AGCAAATTC
551 TTAATTTTGG TAATTTTAT TCTTGGTCTT TTTGTTCTTG GAATCGCCAG
601 CATACTCTAT TACTATTTT CAATGGAAGC AGCAAGTTTA AGTCTCTCCA
651 ATCTTTGGTT TGGATTCTTG CTGGCCTCC TATGTTTCT TGATAATTCA
701 TCCTTTAAAA ATGATGTAAA AGAAGAATCA ACCAAATATT TGCTTCTAAC
751 ATCCATAGTG TTAAGGATAT TGTGCTCTCT GGTGGAGAGA ATTTCTGGCT
801 ATGTCCGTCA TCGGCCCACT TTAATAACCA CAGTTGAATT TCTGGAGCTT
851 GTTGGATTGG CCATTGCCAG CACAACATAT TTGGTGGAGA AGTCTCTGAG
901 TGTCAATTTG CTGTGTGTAG CTCTGGCTAT GCTGATTATT GATCTGAGAA
951 TGAAATCTTT CTTAGCTATT CCAAACCTAG TTATTTTTCG AGTTTGTGTA
1001 TTTTTTTCCT CATTGGAAC TCCCAAAAT CCGATTGCTT TTGCGTGT
1051 TTTTATTTGC CTGATAACTG ATCCTTTCCT TGACATTTAT TTTAGTGGAC
1101 TTTTCAGTAA TGAAAGATGG AAACCTTTT TGTACCGTGG AAGAATTTGC
1151 AGAAGACTTT CAGTCGTTT TGCTGGAATG ATTGAGCTTA CATTTTTAT
1201 TCTTCCGCA TTCAAACCTA GAGACACTCA CCTCTGGTAT TTTGTAATAC
1251 CTGGCTTTTC CATTTTGGGA ATTTTCAGGA TGATTGTGCA TATTATTTT
1301 CTTTAACTC TTTGGGGATT CCATACCAAA TTAATGACT GCCATAAAGT
1351 ATATTTTACT CACAGGACAG ATTACAATAG CCTTGATAGA ATCATGGCAT
1401 CCAAGGGGAT GCGCCATTTT TGCTTGATTT CAGAGCAGTT GGTGTCTTT
1451 AGTCTTCTTG CAACAGCGAT TTTGGGAGCA GTTTCCTGGC AGCCAACAAA
1501 TTGGAATTTT TTGAGCATGT TCCTAATCGT TTGCCATTG GAATCCATGG
1551 CTCATGGGCT CTTCCATGAA TTGGGTAACT GTTTAGGAGG AACATCTGTT
1601 GGATATGCTA TTGTGATTCC CACCAACTTC TGCAGTCCTG ATGGTCAGCC
1651 AACACTGCTT CCCCAGAAC ATGTACAGGA GTTAAATTTG AGGCTACTG
1701 GCATGCTCAA TGCTATCCAA AGATTTTTCG CATATCATAT GATTGAGACC
1751 TATGGATGTG ACTATCCAC AAGTGGACTG TCATTTGATA CTCTGCATT
1801 CAACTAAAA GCTTTCCTCG AACTTCGGAC AGTGGATGGA CCCAGACATG
1851 ATACGTATAT TTTGTATTAC AGTGGGCACA CCCATGGTAC AGGAGAGTGG
1901 GCTCTAGCAG GTGGAGATAC ACTACGCCCT GACACACTTA TAGAATGGTG
1951 GAGAGAAAAG AATGGTTTCT TTTGTTCCCG GCTTATTATC GTATTAGACA
2001 GCGAAAATTC AACCCCTTGG GTGAAAGAAG TGAGGAAAAT TAATGACCAG
2051 TATATTGCAG TGCAAGGAGC AGAGTTGATA AAAACAGTAG ATATTGAAGA
2101 AGCTGACCCG CCACAGCTAG GTGACTTTAC AAAAGACTGG GTAGAATATA
2151 ACTGCACTC CTGTAATAAC ATCTGCTGGA CTGAAAGGG ACGCACAGTG
2201 AAAGCAGTAT ATGGTGTGTC AAAACGGTGG AGTGACTACA CTCTGCATT
2251 GCCAACGGGA AGCGATGTGG CCAAGCACTG GATGTTACAC TTTCTCGTA
2301 TTACATATCC CTAAGTGCAT TTGGCAAAAT GGTATGCGG TCTGAACCTT
2351 TTTGGATCTT GCAAAACTTG TTTAGGTGTC TTGAAAAGAT TAAAAATGAG
2401 TTGGTTTCTT CCTACTGTGC TGGACACAGG ACAAGGCTTC AAACCTGTCA
2451 AATCTTAATT TGGACCCCAA AGCGGGATAT TAATAAGCAC TCATACTACC
2501 AATTATCACT AACTTGCCAT TTTTGTATG CTGATTTTTT ATTTGTGGAA
2551 AATACCTTGC TACTTCTGTA GCTGCTCTCA CTTTGTCTTT TCTTAAGTAA
2601 TTATGGTATA TATAAGGCGT TGGGAAAAAA CATTTTATAA TGAAAGTATG
2651 TAGGGAGTCA AATGCTTACT GTAAATGCAT AAGAGACGTT AAAAAAACA
2701 CTGCACTTTC AGGAATGTTT GCTTATGTC CTGATTAGAA AGAAACAGTT
```

```

2751 GTCTATGCTC TGCAATGGTC AATGATGAAT TACTAATGCC TTATTTTCTA
2801 GGCATATAAT AATAGTTTAG AGAATGTAGA CCAGATAAAT TTGTTTACTG
2851 TTTTAAGAAA ACTACCAGTT TACTTACAGA AGATTCTTTT TTCCAAACAG
2901 TAGGTTTCAT CCAAGACCAT TTGAAGAACT GCAAACCTCT TCTCTTAGAA
2951 AAGAAAGAGG GCAGCCTAAA ATAAACGCAA AATTTGCTTA TACTCCATCA
3001 CATTACAGATG TCTTGGTTGT GACTTATTAC CAGTGTGGCA GAGAACCCTA
3051 GTTACATTTT AGATCAAAAT ATTCTTTATG TAGGTATTGT TAAAAGGCTA
3101 GAGCCTACAA GTTGCTCTTC CATGCGTTGG TCAGGGGGCC CTGAAAACAC
3151 TGGTAATATT AAGAGTCTTT CTCAGGGTAA CTTAATGTTT TCTTAATGAA
3201 CAGTGTTTCC AGCTACAAAT TCTTCCAATA AATTGTCTTC CTTTTTGAAA
3251 AGTACTCTCA TAGAAGAAAT TTAGCAATTT CTCGTTGACT GACTCAGTCT
3301 ATTTTAAGTA TTCAGAAAAG ATTTTGATCC CCATTGAGTT AATGCTCTGC
3351 CTTGAAAATT ATTTTCTGTA TCCTTGTAGG TGATAACATT TTTTCTCTAC
3401 TGAAGGTCAG AGGATAGGAA ACAAGTATTT CTCTTCTGGT ATACATGTAA
3451 TGTATTCTGT AAAAAAGTAT TCATATTGGC AATTTTAGTT AGGCATAATA
3501 TTGTGGTTGT AATTTTAAAT ACTTAGTGTT TTGTCTGATT AAAGCAGGCA
3551 CTGATCAGGG TATCTCCTAA GAGGTAATTC ACTTCTTATT CCTTCCAAT
3601 AATTATTACA TTCTAAATTT TCATCTATGA GAAATAACAA ACAAGAAGGG
3651 AATAGAATTA AATTGGGGTA TAATCTAATC TTCATTGTTT AAATGGTTTG
3701 CCTTCTCACC ATTGAAGCCA TTTTCTTATA GCCTCAGAAA GAGGAAATAA
3751 TGCCTCCACC ATTTTCTACC TGGTGACTTG AAAATTGAAC TTTTAAGTTA
3801 GGAAGAAGTT AGAGTCAGGG AACTTGTATA CCACTATCTA TGCAGCATTG
3851 TTATAGTCTG ATTATTCTCTG TGTTTTGAAT ATGATTTTCC TAATGCTCTA
3901 AATAAAATTT TGTTAAAAAT CAAAAAATAA AAAAAAATAA CTTATCGATA
3951 CCGTCGACCT CGATGATGTC GAC

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 365 bp to 2455 bp; peptide length: 697
 Category: putative protein
 Classification: unset

```

1 MCKSLRYCFS HCLYLAMTRL EEVNVREVMH SSVRYLYGLA RINLLVAICL
51 GLYVRWEKTA NSLILVIFIL GLFVLGSIAS LYYYFSMEAA SLSLSNLWFG
101 FLLGLLCLFD NSSFKNDVKE ESTKYLLLT S IVLRILCSLV ERISGYVRHR
151 PTLTITVEFL ELVGFAIAST TMLVEKSLSV ILLVVALAML IIDLRMKSFL
201 AIPNLVIFAV LFFFSLETP KNPIAFACFF ICLITDPFLD IYFSGLSVTE
251 RWKPFYLRGR ICRRLSVVFA GMIELTFFIL SAFKLDRDTHL WYFVIPGFSI
301 FGIFRMICHI IFLLTLWGFH TKLNDCHKVY FTHRTDYNLS DRIMASKGMR
351 HFCLISEQLV FFSLLATAIL GAVSWQPTNG IFLSMFLIVL PLESMAGLGF
401 HELGNCLGGT SVGYAIVIPT NFCSPDGQPT LLPPEHVQEL NLRSTGMLNA
451 IQRFFAYHMI ETYGCDSYTS GLSFDTLHSL LKAFLELRTV DGPRHDTYIL
501 YYSGHTHTGT EWALAGGDTL RLDTLIEWWR EKNGSFCSRL IIVLDSNST
551 PWVKEVRKIN DQYIAVQGA E LIKTVDIEEA DPPQLGDFTK DWVEYNCNSC
601 NNICWTEKGR TVKAVYGVSK RWSDYTLHLP TGSDVAKHWM LHFPRITYPL
651 VHLANWLCLG NLFWICKTCF RCLKRLKMSW FLPTVLDGTQ GFKLVKS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c1, frame 2

PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*, N = 1,
 Score = 96, P = 0.12

>PIR:A71148 hypothetical protein PH0395 - *Pyrococcus horikoshii*
 Length = 288

HSPs:

Score = 96 (14.4 bits), Expect = 1.3e-01, P = 1.2e-01
 Identities = 59/234 (25%), Positives = 116/234 (49%)

Query: 77 IASILYYFMEAAASLSLNLWFGFLL--GL--LCFLDNSSFKNDVKEESTKYLLTSIV 132
 ++ +LYY F+ A ++ L G+LL + L +L N + V+ + K + ++
 Sbjct: 57 LSLVLYLFAFSALK-TIIFLALGYLLMNSIYELGYLMNDTISRREVEGKVHVRVRLTVF 115

Query: 133 LRILCSLVERISGYVRHRPTLLTTVEFLELVGFIAIASTTMLVEKSLSVILLVVALAMLI 192
 +L +L I YV ++ T+ FL+LVG ++ +L E +L ++ L+ L +
 Sbjct: 116 DSLLIALSRAI--YV-----VIFTLVFLKLVLQYSTQVILAETVTLFLVFLLYDLTPKHV 168

Query: 193 DLRMKSFLAIPNLVIFAVLLFFSSLET-PKNPIAFACFFICLITDPFLDIYFSGLSVTER 251
 M SF + + F +LL F T +N I + FI I F ++ + +
 Sbjct: 169 RTVMSLF-PLKFMKAFVLLLPFIITGTLVENVITLS--FILPIAVRFSQAHYLTACKDN 225

Query: 252 WKPFYLRGRICRRLSVVFAGMIEL-TFFILSAFK-LRDTHLW-YFVIPGFSIFGIFRMIC 308
 P ++ R+ R S+++ + L TF +L +F L +T L ++IP F++ + ++
 Sbjct: 226 -PPRDFKRRV-ERFSMMYLQVTSLSSTFTVLVSEVYLGNTDLLRQYLIP-FAVNVVLILLS 282

Query: 309 HI 310
 ++
 Sbjct: 283 YL 284

Pedant information for DKFZphfbr2_2c1, frame 2

Report for DKFZphfbr2_2c1.2

[LENGTH] 697
 [MW] 79741.46
 [pI] 8.41
 [KW] TRANSMEMBRANE 11
 [KW] LOW_COMPLEXITY 9.76 %

SEQ MCKSLRYCFSHCLYLAMTRLEEVRNREVMHSSVRYLGYLARINLLVAICLGLYVRWEKTA
 SEG
 PRD cccceehhhhhhhhhhhhhhhhhhhhhccceehhhhhhhhhhhhhhhhhhhccccc
 MEMMMMMMMMMMMMMMMMM.....

SEQ NSLILVIFILGLFVLGFIASILYYFMEAAASLSLNLWFGFLLGLLCFLDNSSFKNDVKE
 SEG ..xxxxxxxxxxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
 PRD ccceeeccccchhhccccc
 MEM ..MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ ESTKYLLTSIVLRILCSLVERISGYVRHRPTLLTTVEFLELVGFIAIASTTMLVEKSLSV
 SEGxxxxxxxxxxxxx.....xxxxx
 PRD ccchhhhhhhhhhhhhhhhhhhhhceeeccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
 MEMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....MMM

SEQ ILLVVALAMLIIDLRMKSFLAIPNLVIFAVLLFFSSLET-PKNPIAFACFFICLITDPFLD
 SEG xxxxxxxxxxxxxxxx.....
 PRD hhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
 MEM MMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ IYFSGLSVTERWKPFYLRGRICRRLSVVFAGMIELTFFILSAFKLRDTHLWYFVIPGFSI
 SEG
 PRD eeccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 MEMMMMMMMMMMMMMMMMM.....M

SEQ FGIFRMICHIIIFLLTLWGFHTKLNDCCHKVYFTHRTDYNLDRIMASKGMRHFCLISEQLV
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhcccccceeeccccccccchhhhhhhhhhhhhhhhhhhhhhh
 MEM MMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....MM

SEQ FFSLLATAILGAVSWQPTNGIFLSMFLIVLPLESMAGLFEHLGNCLGGTSVGVAIVIPT
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
 MEM MMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM.....

SEQ NFCSPDGQPTLLPPEHVQELNLRSTGMLNAIQRFAYHMIETYGCDYSTGLSFDTLHSK
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 MEM
 SEG LKAFLELRVDGPRHDYIILYSGHGTGTGEWALAGGDTLRDLTLIEWWREKNGSFC SRL
 SEG
 PRD hhhhhhhhhhhhhhhhhhhhhcccccceeeccccccccceeeccccchhhhhhhhhhhhhhh
 MEM
 SEG IIVLDSENSTPWVKEVRKINDQYIAVQGAELIKTVDIEEADPPQLGDFTKDWVEYNCNSC

```
SEG .....
PRD eeeeeccccccchhhhhccceeeccccceeeccccccccccccceeecccccc
MEM .....

SEQ NNICWTEKGRIVKAVYGVSKRWSYTLHLPTGSDVAKHWMHLHFPRITYPLVHLANWLCGL
SEG .....
PRD ceeeeccccceeeccccceeeccccchhhhhccceccccchhhhhhhhhcc
MEM .....

SEQ NLFWICKTCFRCLKRLKMSWFLPTVLDTGQGFKLKVS
SEG .....
PRD eeeeehhhhhhhhhhhhccceeecccccccccc
MEM .....
```

(No Prosite data available for DKF2phfbr2_2c1.2)

(No Pfam data available for DKF2phfbr2_2c1.2)

DKF2phfbr2_2c17

group: signal transduction

DKF2phfbr2_2c17.3 encodes a novel 446 amino acid protein with similarity to yeast YMR131c and mammalian Retinoblastoma-binding protein RbAp46

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to YMR131c and retinoblastoma-binding protein RbAp46

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2248 bp

Poly A stretch at pos. 2230, polyadenylation signal at pos. 2200

```
1 TGGGGAAGAT GGCGGCGCGC AAGGGTCGGC GTCGCACGTG TGAAACCGGG
51 GAACCCATGG AAGCCGAGTC CGGCGACACA AGTTCGAGG GCCCGGCCCA
101 GGCTCTACCTG CCCGGCCGGG GGCCGCCGCT ACGCGAAGGG GAGGAGCTGG
151 TCATGGACGA GGAGGCCTAT GTGCTCTACC ACCGAGCGCA GACTGGCGCC
201 CCCTGTCTCA GCTTTGACAT AGTCCGGGAT CACCTGGGAG ACAACCGGAC
251 AGAGCTTCTT CTTACACTTT ACTTGTGTGC TGGGACCCAG GCTGAGAGCG
301 CCCAGAGCAA CAGACTGATG ATGCTTCGGA TGCACAATCT GCATGGGACA
351 AAGCCCCAC CCTCAGAGGG CAGTGATGAA GAAGAAGAGG AGGAAGATGA
401 AGAGGATGAA GAAGAGCGGA AACCTCAGCT GGAGCTGGCC ATGGTGCCCC
451 ACTATGGTGG CATCAACCGA GTTCGGGTGT CATGGCTGGG TGAAGAGCCT
501 GTGGCTGGGG TGTGGTCAGA GAAGGCCAG GTGGAGGTGT TTGCGCTGCG
551 GCGGCTTCTG CAGGTGGTGG AGGAGCCCCA GGCCCTGGCA GCCTTCCTCC
601 GGGATGAGCA GGCCCAAATG AAGCCATCT TCTCCTTCGC TGGACACATG
651 GGCAGGGGCT TTGCCCTTGA CTGGTCCCCC CGGGTGACCG GTCGCCTGCT
701 GACCGGTGAC TGTCAAAGA ACATCCACCT CTGGACACCT ACGGACGGCG
751 GCTCTGGCA CGTGGACCAG CGGCCATTG TGGGCCACAC ACGCTCTGTG
801 GAGGACCTGC AGTGGTCACC GACTGAGAAC ACGGTGTTTG CCTCTGTCTC
851 AGCTGACGCC TCCATCCGCA TCTGGGACAT CCGGGCAGCC CCCAGCAAGG
901 CCTGCATGCT CACCACAGTC ACCGCCCATG ATGGGGAGCT CAATGTATC
951 AGCTGGAGCC GCCGGGAGCC CTCTCTGCTC AGTGGCGGGG ATGATGGGGC
1001 CCTCAAGATC TGGGACCTTC GGCAGTTCAA GTCTGGTTCC CCAGTGGCCA
1051 CCTTCAAGCA GCACGTGGCC CCCGTGACCT CCGTCGAGTG GCACCCCCAG
1101 GACAGCGGGG TCTTTGCAGC CTCGGGTGCA GACCACAGA TCACACAGTG
1151 GGACCTGGCA GTGGAGCGGG ACCCTGAGGC GGGCGACGTG GAGGCCGACC
1201 CCGGACTGGC CGACCTCCCG CAGCAGCTGC TGTTCGTGCA CCAGGGCGAG
1251 ACCGAGCTGA AGGAGCTGCA CTGGCACCCG CAGTGCCAG GGCTCCTGGT
1301 CAGCACGGCG CTGTCAGGCT TCACCATCTT CCGCACCATC AGCGTCTGAG
1351 GCGTCCCACT GGCTCTGATC TTGCTTCCTG CTTGGAAACT GAAGTCGAAT
1401 TGGGCTCCCC TGGAAAGGGT TCATTCAAGT CTGTTGACTG AGACTGGCCG
1451 GCCTGTGGGC TGCCGTGATG GATTCTGTTT GACGTATTGT TCTCTAGAAG
1501 GCCTGGCTCT GATCCAGTGA CCCCTCTCAC CAAAGAACTC GGTTTAACCA
1551 GGGCTCTGTA AGACCACTCC CACCCAGAGA CTTGTGTGGC CTGGTGTGGC
1601 CTGTGTGTCG GATTCCCTCC TGTCAGCTGT GACCCATTG ACCTGTGTCC
1651 CCAGAACCCA GTTTTTTGTG TGTTTGTTG AGACGGAGTC TTGGTCTGTC
1701 GCCCAGGCTG GAGTGCACTA GCACGATCTT GGCTCACTGC AACCTCCGCC
1751 TCCTGGGTTA AAGTGATTCT CTCAGCTCAG TCTCCAGGT AGCTGGGATT
1801 ACAGGCATGT GCCACCACAC CCCGTTAATT TTTGTATTTT TAGTAGAGAC
1851 GGGGTTTCAC CATGTTGGCC AGGCTGGTCT CAAATCTTGT ATCTCAAGTG
1901 ATCTGTCCGC CCCGGCTCC CAGAGTGCTG GGTGGGATT ACAGGCGTGA
1951 GCCACCGGCT CCGGCTCAGG ACCCAGTTT GGCTGCTGGT TCCCAGCAGG
2001 GGACTCGGGG GATATACAGT GGCTGCACCA AATTGGAGGT GTGGGTTCTT
2051 CCAACACAAT TTGCTTCTGC CCGTTGTCTT CCTGCCAGCT GGGTTTGCC
2101 AGGATTTCTC CGTGTGGGG CTACATGCGA CCCTCTCCCC TCCTCCCTGA
2151 CTTTAGAGGC TGGTGTCTGT TCGGGAGGAA GGTCAGGGCT CCTGAGCAGC
2201 AATAAAGGAC CAGGAAGAGG CCTGAGGTGG AAAAAAAAAA AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 9 bp to 1346 bp; peptide length: 446
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: WD_REPEATS (323-338)

```

1 MAARKGRRRT CETGEPMEAE SGDTSSSEGA QVYLPGRGPP LREGHEELVMD
51 EEAYVLYHRA QTGAPCLSFD IVRDHLGDNR TELPLTLYLC AGTQAESAQS
101 NRLMMLRMHN LHGTPKPPSE GSDEEEED EDEEEKRPQ LELAMVPHYG
151 GINRVRVSWL GEEPVAGVWS EKGQVEVFAL RRLQVVEEP QALAAFLRDE
201 QAQMKPIFSF AGHMGEGFAL DWSPRVTGRL LTGDCQKNIH LWTPTDGGSW
251 HVDQRPFVGH TRSVEDLQWS PTENTVFASC SADASIRIWD IRAAPSKACM
301 LTTVTAHDGD VNVISWSRRE PFLSGGDDG ALKIWDLRQF KSGSPVATFK
351 QHVAPVTSVE WHPQDSGVFA ASGADHQITQ WDLAVERDPE AGDVEADPGL
401 ADLPQQLLFV HQGETELKEL HWHPQCPGLL VSTALSGFTI FRTISV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2c17, frame 3

TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence., N = 1, Score = 910, P = 2.7e-91

PIR:S53061 hypothetical protein YMR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 691, P = 4.3e-68

PIR:I49367 retinoblastoma-binding protein mRbAp46 - mouse, N = 1, Score = 338, P = 1.1e-30

PIR:I39181 retinoblastoma-binding protein RbAp46 - human, N = 1, Score = 338, P = 1.1e-30

>TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein"; Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence.
 Length = 469

HSPs:

Score = 910 (136.5 bits), Expect = 2.7e-91, P = 2.7e-91
 Identities = 195/442 (44%), Positives = 259/442 (58%)

```

Query: 18 EAESGDTSSSEGAQVYLPGRGPPLREGEELVMDEEAYVLYHRAQTGAPCLSFDIVRDHLG 77
      EA S + S P +V+ PG L +GEEL D AY H G PCLSFDI+ D LG
Sbjct: 18 EASSSEIPSI-PTRVWQPGVDT-LEDGEELQCDFSAYNSLGHFHVGPCLSFIDILGDKLG 75

Query: 78 DNRTELPLTLYLCAGTQAESAQSNRLMMLRMHNLHGTPK---PPSEGSDEEEED- 133
      NRTE P TLY+ AGTQAE A N + + ++ N+ G + P + G+ E+E+E+DE+D
Sbjct: 76 LNRTEFPHTLYMVAGTQAEKAAHNSIGLFKITNVSGKRRDVVPKTFNGEDEDEDEDDDS 135

Query: 134 -----EEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKQVEVFALRRLQ 185
      E + P.+++ V H+G +NR+R + W++ G V+V+ + L
Sbjct: 136 DSDDDGDGEASKTPNIQVRRVAHGCNVNRIRAMPQNSH-ICVSWADSGHVQVWDMSSHLN 194

Query: 186 VVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIHLWTPT 245
      + E + P+ +F+GH EG+A+DWSP GRLL+GDC+ IHLW P
Sbjct: 195 ALAESETEGKDGTSFVLNQAPLVNFGSGHKDEGYAIDWSPATAGRLSGDCKSMIHLWEPA 254

Query: 246 DGGSWHVDQRPFVGHTRSVEDLQWSPENTVFASCADASIRIWDIRAAPSKACMLTTVT 305
      G SW VD PF GHT SVEDLQWSP E VFASCS D S+ +WDIR S A +
Sbjct: 255 SG-SWAVDPIPFAGHTASVEDLQWSPAENNVFASCSVDGSAVAVWDIRLGKSPAL---SFK 310

Query: 306 AHDGDNVVISWSRREPFL-L-SGGDDGALKIWDLRQFKSGSPV-ATFKQHVAPVTSVEWHP 363
      AH+ DVNVISW+R +L SG DDG I DLR K G V A F+ H P+TS+EW
Sbjct: 311 AHNADVNVISWNRLASCLASGSDGTFSIRDLRLIKGGDAVVAHFEYHKHPITSIEWSA 370

```


Query: 364 QDSGVFAASGADHQITQWDLAVERDPE-----AGDVEADPGLADLPQQLLFVHQGETEL 417
 ++ A + D+Q+T WDL++E+D E A E DLP QLLFVHQG+ +L
 Sbjct: 371 HEASTLAVTSGDNQLTIWDLSEKDEEEAEFNAQTKELVNTPDQLPQQLLFVHQGQKDL 430

Query: 418 KELHWHFQCPGLLVSTALSGFTIFRTISV 446
 KELHWH Q PG+++STA GF I ++
 Sbjct: 431 KELHWHNQIPGMIISTAGDGFNILMPYNI 459

Pedant information for DKFZphfbr2_2c17, frame 3

Report for DKFZphfbr2_2c17.3.

[LENGTH] 446
 [MW] 49447.38
 [pI] 4.82
 [HOMOL] TREMBL:AC005917_14 gene: "F3P11.14"; product: "putative WD-40 repeat protein";
 Arabidopsis thaliana chromosome II BAC F3P11 genomic sequence, complete sequence. 1e-90
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YMR131c] 4e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 4e-15
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 2e-13
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-11
 [FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 4e-09
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 5e-09
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 5e-09
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 6e-09
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 5e-08
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-08
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YLR429w] 3e-07
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 3e-06
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 4e-06
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 4e-06
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-06
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-05
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c] 2e-05
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w] 2e-05
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 2e-05
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL011w] 3e-05
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YOR212w] 5e-05
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 5e-05
 [BLOCKS] BL00678
 [SCOP] d2trcb_2.51.3.1.1 Transducin (heterotrimeric G protein), gamm 5e-29
 [PIRKW] plasma 6e-07
 [PIRKW] duplication 4e-12
 [PIRKW] hormone 6e-07
 [PIRKW] transmembrane protein 1e-07
 [PIRKW] stomach 6e-07
 [PIRKW] actin binding 1e-07
 [PIRKW] leucine zipper 1e-07
 [PIRKW] signal transduction 2e-06
 [PIRKW] heterotrimer 2e-06
 [PIRKW] peripheral membrane protein 6e-07
 [PIRKW] GTP binding 2e-06
 [SUPFAM] WD repeat homology 1e-63
 [SUPFAM] yeast coatamer complex alpha chain 1e-07
 [SUPFAM] GTP-binding regulatory protein beta chain 4e-07
 [SUPFAM] PRL1 protein 8e-09

[SUPFAM] MS11 protein 4e-12
 [SUPFAM] coatomer complex beta' chain 1e-09
 [PROSITE] WD_REPEATS 1
 [PFAM] WD domain, G-beta repeats
 [KW] All_Beta
 [KW] 3D
 [KW] LOW_COMPLEXITY 3.14 %

```

SEQ  MAARKGRRRTCETGEPMEAESGDTSSGPAQVYLPGRGPPLREGSELVMDDEEAYVLYHRA
SEG  .....
lgotB .....

SEQ  QTGAPCLSFDIVRDHLGDNRTPLTLYLCAQTQAESAQSNRLMMLRMHNLHGKPPPE
SEG  .....
lgotB .....

SEQ  GSDEEEEEDEEDEERKPQLELAMVPHYGGINRVRVSWLGEEPVAGVWSEKQVEVFAL
SEG  ..XXXXXXXXXXXXXXXXX.....
lgotB .....

SEQ  RRLQVVEEPQALAAFLRDEQAQMKPIFSFAGHMGEGFALDWSPRVTGRLLTGDCQKNIH
SEG  .....
lgotB .....EEECCECECECEETTT-TCEEEEEEETTEEE

SEQ  LWTPTDGGSWHVDQRPFVGHTRSVEDLQWSPTENTVFASCSADASIRIWDIRAAPSKACM
SEG  .....
lgotB EEEETTT---CEEEEEECECECEEEETTTCE-EEEEETTTTEEEETTT--TEEEE

SEQ  LTTVTAHDGDVNVISWSRREPFLSGGDDGALKIWDLRQFKSGSPVATFKQHVPVTSVE
SEG  .....
lgotB EECBTTBTCCEEEEEEETTTTEEEEEEETTTTEEEEEE.....

SEQ  WHPDQSGVFAASGADHQITQWDLAVERDPEAGDVEADPGLADLPQQLLFVHQGETELKEL
SEG  .....
lgotB .....

SEQ  HWHPQCPGLLVSTALSGFTIFRTISV
SEG  .....
lgotB .....

```

Prosites for DKF2phfbr2_2c17.3

PS00678 323->338 WD_REPEATS PDOC00574

Pfam for DKF2phfbr2_2c17.3

HMM_NAME WD domain, G-beta repeats

HMM *MrGHnnWVWCVaFSPDGGrWFIvSGSWDgTCRLWD*
 ++GH+ V ++ +SP + +++S S D ++R+WD

Query 257 FVGHTRSVEDLQWSPTENTVFASCSADASIRIWD 290

24.88 304 336 1 34 dkf2phfbr2_2c17.3 similarity to YMR131c and retinoblastoma-binding protein RbAp46

Alignment to HMM consensus:
 Query *MrGHnnWVWCVaFSPDGGrWFIvSGSWDgTCRLWD*
 + H+++V+ +++S + ++SG++DG +++WD

dkf2phfbr2 304 VTAHDGDVNVISWSRREPFLSGGDDGALKIWD 336

DKFZphfbr2_2c18

group: brain associated

DKFZphfbr2_2c18 encodes a novel 302 amino acid protein with weak similarity to cyclin-dependent kinase p130-PITSLRE.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

weak similarity to cyclin-dependent kinase p130-PITSLRE

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2835 bp

Poly A stretch at pos. 2817, polyadenylation signal at pos. 2796

```
1 TGGGGCGGAC GCGAGGGAG TCCAGAGCCT TGAGCCCGGT GCTCCTCCCT
51 CGCGCAGCGG TGGCTCTGCG GCCGCTGGAG TAAACACTGC CTTTGTTCCT
101 TAGCGCCTCG TCTTTCGTCG CCCCGTGCCC TCACGCCGCC GGGCTCTGGC
151 CGGCCCGCCC TCGGTCTCTG AACCCCATTT CGGCTCGTGC CGTGCGGATG
201 CAGCTGCCGG GCCTGGGTTT GGGCATTGAG CGGGAGGAGG AGGAGGAGCG
251 GCGGCGCCTG GCGGCGCATG GATGGGGAAC TGCTGCTGGA CGCAGTGCTT
301 CGGACTGCCT CGCAAGGAAG CGGGGCGGCT GCAGCGAGTA GCGGCGGCGG
351 GAGGATCCAA GTATTTTAGA ACATGCTCAA GAGGTGAGCA CTTGACAATA
401 GAGTTTGAGA ATCTAGTAGA AAGTGATGAA GGGGAGAGCC CAGGAAGCAG
451 TCATAGGCCT CTTACTGAGG AAGAAATGTT TGACCTAAGA GAAAGGCATT
501 ATGATTCCAT TGCCGAAAAA CAAAAAGATC TTGATGAGAA AATTCAAAAA
551 GAGTTAGCCT TACAAGAAGA GAAGTTAAGA CTAGAAGAAG AAGCTTTATA
601 CGCTGCACAG CGTGAAGCAG CCAGGGCAGC AAAGCAGCGA AAGCTCTTGG
651 AGCAAGAAAG GCAGAGAATT GTGCAGCAAT ATCATCCTTC CAACAATGGA
701 GAATATCAAA GTTCAGGACC AGAAGATGAC TTCGAATCTT GTTTGAGAAA
751 TATGAAGTCA CAGTATGAAG TTTTTCGAAG TAGTAGACTC TCATCAGATG
801 CTACAGTTTT GACACCAAAAT ACAGAAAGCA GTTGTGATTT AATGACCAAA
851 ACTAAATCAA CTAGTGGAAG TGACGACAGC ACATCCTTAG ATCTAGAGTG
901 GGAAGATGAA GAAGGAATGA ATAGAATGCT TCCAATGAGA GAACGTTCCA
951 AAACAGAGGA AGACATTCTA CGGGCAGCAC TTAAGTATAG CAACAAGAAG
1001 ACTGGAAGTA ATCCTACATC AGCCTCTGAT GATTCCAATG GGCTGGAGTG
1051 GGAAATGAT TTTGTTAGTG CCGAAATGGA TGATAATGGA AATTCCGAGT
1101 ATTCTGGATT TGTAATCCTT GTATTAGAAC TGTCTGATTC TGGCATAAGG
1151 CATTCTGACA CAGATCAACA GACTCGATAG GGTAAATTTG TGTGACCTTG
1201 TTTATCAGTT ATGACCAAAAT GTTAAAAACC AACTAGAATG TATAAGTGAT
1251 TGTGCTTAGC CTTTTGTGTA GGGAGATGTG TAAGAAACCA TGCTGTAAT
1301 GCTTATTTTA TTACAAAGGA GTAGGGATGA TAGGATCTGA ATTGATACAG
1351 AATTAAAGTG AATTTTCATCA TCTGCCCTCT GCTTTTCAAG ACCAATTTAA
1401 TGGTCTGTGC ATGTTACTGA TTAATTTTAC TTTGCTTTGT CTTTATAGCA
1451 TTTCTGTTTA CTATGGTAGA TTTCCACTTT CAATTTTAA AATTAATTTT
1501 ACTTTGAATG ATTTATGAAG CCTATTTTCT TGTCTAATA TGAATATATT
1551 AAGACTTTTT TGTTAATTTT CAGCCGATGT GAAGGAAGCA TGAGGAGGGA
1601 TCGTCAGACT CAGATTTAGA ATAGTGTTCG CGTTTCCAGC ATTATTTATT
1651 TCTATGACTT CTTTGGATTT TATTATCTAA TAGTAAGTAC AGTTGATGTG
1701 GGTAGATGAC TCTAAGAAAT GCTGAAGTAT CGGCATTACA TGTGTTTATT
1751 TACATGTCCT AGTTTGATAA TGTTGATTCA ATCTGAACAA AAGATAATAT
1801 AAAAAATAAC CTTCAGAGTT TGGACATTTC AAGTTGGTAA TAATAAAAAA
1851 TAATATTTAA GAAGATATAT ATATATATAT ATTTAGTTTT TTCCACTTCA
1901 TTTTACATGC CACTATATTG ACTTTAATTG ATATACAGTA TTAAGTTTTT
1951 AGGTGCCATT ATTTTAAAAA AATTCTATAT TTCCAATGAA CGATGTTAGA
2001 TTTTACACAG AACATATTCT CTGCATGATT TCAGAAAAGA AAATCTAAAA
2051 AGGTAATACG GGTATTTCAA ATAAATCCTT TTCTGGTATG AAAGGCTCCA
2101 TTGATTTTAT TAAGCCTTCC TTTACCTTGT AGTACAAGGT GCTTTAATGG
2151 ATGTGGCTCT TCAATATTTT TATCTATAAC TGCATTTTGT CTAGACAAT
2201 TACTGTTCTT TTCTCTAAAA TGTATATGTC AATTTACAAG GCCAGGGATA
2251 GAAACACATC CATAAATGCT TTCCTTGATT TTGCTGAGGA TTTGGTATGA
2301 TTTTAGTAAG CAAACTGTTT TTTGGTTTTT CCTTAATGTT TTTAATTTTT
2351 TTTCTCTTGG CAACAATGAC GGTGCATGTT CTTATAAATA TAGGAAGGTC
2401 CAGATATAAA TAGTAACCTA AAGTTCCTGC TGTGCTTAAA AAAAAAATC
2451 ATGTGGCTCT TCAATATTTT GAACTGCTAA GCAATGCATC CTGTAGTTTT
2501 ATCTCCTTTT TTATGTCATA GAAATTAATA TGATACTTTA AATATGTAAA
2551 TATAATACAT TGGTAATGCT ATTATTTATA TCTGTCTTAA CATAATTTAA
2601 GTTGTAGCTG TGTCTGGGAA ATATTTTAA GGTAACTCTAT ATTCACATTG
2651 CCTGTGTTAA TGCTTTTAA GGTGTTGATA CATCAGATGT ATATTTTGG
```

2701 TTTGGCATAA GCTACGATTG TAATTTTCT TGGCTTTTGT TTCATAAAGA
2751 ATTTTTTGAA GGAATGGTAA CAAATGGTAA TTTACAAATG GTTGTGAATA
2801 AACACATTTT TACACTTAA AAAAAAAAAA AAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 272 bp to 1177 bp; peptide length: 302
Category: similarity to known protein

1	MGNCCTWTCF	GLLRKEAGRL	QRVGGGGGSI	YFRTCSRGEH	LTIEFENLVE
51	SDEGSEPGSS	HAQLETEEIV	DLRRHRYSKI	AEKQGLDLDEK	IQKELALQEE
101	KRLREEALEY	AAPLREAAAA	KQKRLLEQER	QRIVQOYHPS	NNQEYQSSGP
151	EDDFESCRLN	MKQSYVFRFS	SRLLSDATVL	PTNTSSCDL	MKTKTSTSGN
201	DDSTSLDELE	EDEEGMNRFL	PMRRESATPE	DLTLAALYKS	NKKTGSNPTS
251	ASDDSNGLWE	ENDFVSAEMD	DNGNSEYSGF	VNPVLELSDS	GIRHSDTPDQ
301	TR				

BLASTP hits

Entry A55817 from database PIR:
cyclin-dependent kinase p130-PITSLRE - mouse
Length = 783
Score = 123 (43.3 bits), Expect = 0.00013, P = 0.00013
Identities = 53/197 (26%), Positives = 96/197 (48%)

Alert BLASTP hits for DKFZphfbr2 2c18, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 2c18, frame 2

Report for DKF2phfbz2 2c18.2

[LENGTH]	302	
[MW]	34281.39	
[pI]	4.73	
[PROSITE]	MYRISTYL	5
[PROSITE]	CK2_PHOSPHO_SITE	12
[PROSITE]	TYR_PHOSPHO_SITE	2
[PROSITE]	PKC_PHOSPHO_SITE	3
[KW]	All_Alpha	
[KW]	LOW COMPLEXITY	13.58 %
[KW]	COILED COIL	13.58 %

```

SEQ      MGNCWCCTQCFGLLRKEAGRLQRVGGGGGSKYFRTCSRGEHLTIEFENLVESDEGESPGSS
SEG      .....xxxxx.....
PRD      cccccccchhhhhhhhhheecccccccceeecccccchhhhhhhhhccccc
COILS    .....

SEQ      HRPLTEEEIVDLRERHYSIAEKQKDLDEKIQKELALQEEKLRLEEALYAAQREAAARA
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccchhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      KQRKLEQERQRIVQYHPSPNNGEYQSSGPEDDDFESCLRNMKSQYEVFRSSRLSSDATVL
SEG      xxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhccccc(ccccccccccccchhhhhhhhhhhhhheeeeccccceeee
COILS    CCCCCCCCCC

```

```

SEQ      TPNTESSCDLMTTKTSTSGNDSTSLDLEWEDEEGMNRMLPMRERSKTEEDILRAALKYS
SEG      .....
PRD      cccccccccccccccccccccchhhhhhhccccccchhhhhhhcchhhhhhhhhhhc
COILS    .....

SEQ      NKKTGSNPTSASDDSNLEWENDFVSAEMDDNGNSEYSGFVNPVLELSDSGIRHSDTDQQ
SEG      .....
PRD      cccccccccccccccccccccceeeccccccccccccccccceeecccccccccccccc
COILS    .....

SEQ      TR
SEG      ..
PRD      cc
COILS    ..

```

Prosites for DKFZphfbr2_2c18.2

PS00005	60->63	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	65->69	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	163->167	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	198->202	CK2_PHOSPHO_SITE	PDOC00006
PS00006	204->208	CK2_PHOSPHO_SITE	PDOC00006
PS00006	226->230	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	295->299	CK2_PHOSPHO_SITE	PDOC00006
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00007	103->111	TYR_PHOSPHO_SITE	PDOC00007
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	199->205	MYRISTYL	PDOC00008
PS00008	245->251	MYRISTYL	PDOC00008
PS00008	291->297	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_2c18.2)

DKFZphfbr2_2d15

group: differentiation/development

DKFZphfbr2_2d15 encodes a novel 438 amino acid protein similarity to Mus musculus testis-specific Y-encoded-like protein (Tspyl1).

The TSPY genes are arranged in clusters on the Y chromosome of many mammalian species. TSPY is believed to function in early spermatogenesis and is a candidate for GBY, the putative gonadoblastoma-inducing gene on the Y. The novel protein is a new member of the TSPY-SET-NAP1L1 family, which represents proteins closely related to TSPY. Therefore, the new protein seems to be involved in early spermatogenesis.

The new protein can find application in modulating early spermatogenesis.

strong similarity to testis-specific Y-encoded-like protein

complete cDNA, complete cds, EST hits
localisation: primer B does not match perfect

Sequenced by Qiagen

Locus: /map="729.2 cR from top of Chr6 linkage group"

Insert length: 3229 bp

Poly A stretch at pos. 3206, polyadenylation signal at pos. 3184

```
1 GGAGACTGTA GGGTGGGCGG TCGGAGCGGC GGTAGCTCC CAGTTCGGCC
51 TCTGAGGAAA ACGGGCGTTC GCCTGCGGTT GGTCCGACTG TTAGCAACAT
101 GAGCGGCGTG GATGGGGTCA AGAGGACCAC TCCCCCCTCA ACCCAGAGCA
151 TCATTATTTC TGACCAAGTC CCGAGCGACC AGGACGCACA CCAGTACCTG
201 AGGCTCCGCG ACCAAAGCGA GGCGACACAG GTGATGGCGG AGCCGGGTGA
251 GGGAGGGCTCG GAGACCGTCG CGCTCCCCGC TTCACCGCCT TCAGAGGAGG
301 GGGGCGGTACC CCAGGATCCC GCGGGCCGTG GCGGTACTCC CCAGATCCGA
351 GTTGTGGGGG GTCGCGGTCA TGTGGCGATC AAAGCCGGGC AGGAAGAGGG
401 CCAGCCTCCC GCCGAAGGCC TGGCAGCCGC TTCTGTGGTG ATGGCAGCCG
451 ACCGCAGCCT GAAAAGGGC GTTCAGGGTG GAGAGAAGGC CCTAGAAATC
501 TGTGGCGCCC AGAGATCCGC GTCTGAGCTG ACGGCGGGGG CGGAGGCTGA
551 GCGGAGGAGG GTGAAGACAG GAAAGTGC GCACCTCTCA GCAGCCGTGG
601 CTGAGAGGGA GAGCGCTGAG GTGGTGGTGA AGGAAGGCCT GCGGAGAAAG
651 GAGGTAATGT AGGAGCAGAT GGAGGTAGAG GAGCAGCCGC CAGAAGGTGA
701 AGAATAGAAA GTGGCGGAGG AGGATAGAT GGAGGAGGAG GCGAGGGAGG
751 AAGAAGGGCC CTGGCCTTTG CATGAGGCTC TCCGCATGGA CCCTCTGGAG
801 GCCATCCAGC TGGAACTGGA CACTGTGAAT GCTCAGGCCG ACAGGGCCTT
851 CCAACAGCTG GAGCACAAGT TTGGGCGGAT GCGTCGACAC TACCTGGAGC
901 GGAGGAACCTA CATCATTCAG AATATCCCGG GCTTCTGGAT GACTGCTTTT
951 CGAAACCACC CCCAGTTGTC CGCCATGATT AGGGGCCAAG ATGCAGAGAT
1001 GTTAAGGTAC ATAACCAATT TAGAGGTGAA GGAACCTAGA CACCTAGAAA
1051 CCGGTGTCAA GTTCAAGTTC TTCTTTAGAA GAAACCCCTA CTTCAGAAAC
1101 AAGCTGATTG TCAAGGAATA TGAGGTAAGA TCCTCCGGCC GAGTGGTGTC
1151 TCTTTCTACT CCAATTATAT GGCGCAGGGG GCATGAACCC CAGTCCTTCA
1201 TTCCAGAGAA CCAAGACCTC ATCTGCAGCT TCTTCACTTG GTTTTCAGAC
1251 ATACAGCCTC CAGAGTCCGA CAAATTTGCT GAGATTATTA AAGAGGATCT
1301 GTGGCCAAAT CCACTGCAAT ACTACCTGTT GCGTGAAGGA GTCCGTAGAG
1351 CCCGACGTCG CCCGCTAAGG GAGCCTGTAG AGATCCCCAG GCCCTTTGGG
1401 TTCCAGTCTG GTTAACATTT GCCCTTGGGA ATACTCTTGC ACAAGGTCTC
1451 CTACCACCTT CTGCTGGACC TGTGCTTGGG CATCAGCAAT GAGTATGCCT
1501 TCTATTGTGC TTTGTTTTTG CTGACTTTTC TGCACCTGTG TTCCTTTGGA
1551 TATTCACTTC TCTCAACCTC AAGATTGAGA CGGTGGTGGG TATGCTTCTC
1601 CACTTCCATA TGACCTTCAT GCTGTTCTGG AATATCACAT GCTACGAGGT
1651 CATCCTTCACT ACTACTTGTA AGCCAAGCAA ATGATACTGT AGATTGTACT
1701 GCCTTTATCT GCACTGCTTG GACCTGTTT ATTCCAGGG CCTCTGAAC
1751 GGTGCTGTC ACTTGGATTT CTAGCTTTGG GAGCCTGTTC CACCTACTCA
1801 GCTCTGCATT GAGCAGTATG GGCACATGCC CTGTGGACAG TTACTGGACG
1851 TTAATGAACCT CAGAGGAGAA AAGCAGTGAG CCACTGTGTC TGTGTGATTT
1901 ATGGTACTTC ATTGCTCTTC CTTCACTCTC AGTCACTTTC TATTGCTACC
1951 TGCCCTACAT TGGCTCCTGC CAAGGTCCCT CTCTCTCCCT GTTTTCTTTT
2001 TTTTTTTTTT TTTTTTTTTT TTTTGAGACG GAGGACGGAG TCTTGCTCTG
2051 TCGCCAGGTT TGGAGTGCAG TGGCGCGATC TCGGCTCACT GCAACCTCCA
2101 CCTCCCGGGT TCAAGCGATT CTCCTGCCIC AGCCTCCGGA GTAGCTGGGA
2151 CTACAGGCGC GCGCCGCCAC GCCCGGCTAA TTTTATATAT TTTAGTAGAG
2201 ACGGGGTTCT ACCATGCTGG CCAGGCTGGT CTCGAACCCC GACCTCGTGA
2251 TCCGCCCTCC TTAGCCTCCC AATCCTCTCT TAAAAAGTGT ATAGCTCAGA
2301 AATATTTGTA AAAGCAAGGT TTTTATTTCA TTTTGGCTCT GTCAATTTCA
2351 GAGGCAAAAG AGTTGGCCTG TAAATAGAG TGCTAGAGCT CTTACGCCCC
2401 TCCCTTCTCT CCCAACTTCC TACTTCTTAG CCCTTTTATC AACTCCTAGA
2451 ATAGTTAAAG AGAGACACAT CTAGATGGGA TGAAAGGTGC CCTAAGCAGG
```

```

2501 AGAACTGAA CAAAAGGCTA GAGGCATGGG CCAGGTAAAA ATTGGGCCTA
2551 GAGTGAAGAC TGTGCTGCCG TTAAGAGCTT TCGAGGAAGG AGTACTTACT
2601 CCCCAATGAT GATGAATGGA GAAATACTTT TCAGGGAGAA TTGAAGGGGT
2651 TAAAGTGTTA AATATGTTGC CTAGACAAGG GTTCTTTAAA GAAAGACAGC
2701 GCAACTTTGA ATGCTTTCTT ACTTGTTTGG TGACCTAATT TATGTGGAAG
2751 ATTGTTATTT CATTAGGATT TAGTAAATTT TTTTCTCTG ATTCTAACT
2801 TATTGTGAAA ATTGAGCTGT ACAGATATTC TTTTGATTTC AATTGGGAAC
2851 ATTTGGAAGA ACAACAGTCT TACTTGCCTG TACAATATAG AGACATATGA
2901 ATAGTCATAA CAGTTTTCAA CTTGTTCTTG TTTCTGTAA ACTATATTCC
2951 TAGAAACATA GTTTGAACAA CTTGGTCTTT GTTAGGCTTG TCAAATTGCC
3001 TTCATGGAAA AATAATCTAC AAAAGTATGG TTTAATTGAT TGTCTTACAT
3051 GATAATTTTC CCTGGCAACA ACTTAGTAAG TGATATATCT TTTTCTCTAA
3101 ATTGCTTAAA TACTGTGAAA TTGCTCTGAC AAATTGGAAG TGTACCATTG
3151 GCATATTGTG CTTCTTTTTT ATGCATGATG GTAAATAAAA AGCATGTTGT
3201 TCTGCTAAGA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry AF042181 from database EMBLNEW:
Homo sapiens testis-specific Y-encoded-like protein (TSPYL) mRNA,
partial cds.

Score = 3411, P = 6.9e-148, identities = 685/687

Entry HS938343 from database EMBL:
human STS WI-11947.

Score = 1195, P = 2.1e-46, identities = 273/299

Medline entries

98399864:

Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1 family

Peptide information for frame 3

ORF from 99 bp to 1412 bp; peptide length: 438

Category: strong similarity to known protein

Classification: Differentiation/Development

```

1 MSLGLDGVKRT TPLQTHSIII SDQVPSDODA HQYLRLRDQS EATQVMAEPG
51 EGGSETVALP PSPPSEEGGV PQDPAGRGGT PQIRVVGGRG HVAIKAGQEE
101 GQPPAEGLAA ASVVMADRS LKKGVGQGEK ALEICGAQRS ASELTAGAEA
151 EAEEVKTKGK ATVSAVAER ESAEVVVKEG LAEKEVMEEQ MEVEEQPPEG
201 EEEVVAEEDR LEEEAAREEG PWPLHEALRM DPLEAIQLEL DTVNAQADRA
251 FQQLHKKFGR MRRHYLERRN YIIQNIPGFW MTAERNHPQL SAMIRGQDAE
301 MLRYITNLEV KELRHPRGTC KFKFFFRNP YFRNKLIVKE YEVRSSGRVV
351 SLSTPTIWRG GHEPQSFIRR NQDLICSFFT WFSDHSLPES DKIAEIIKED
401 LWPNPQQYYL LREGVRRARR RPLREPVEIP RPFQFQSG

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d15, frame 3

TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds., N = 1, Score = 1202, P = 3.1e-122

TREMBL:AB018264_1 gene: "KIAA0721"; product: "KIAA0721 protein"; Homo sapiens mRNA for KIAA0721 protein, partial cds., N = 1, Score = 798, P = 2e-79

TREMBL:AB015345_1 gene: "HRIHFB2216"; Homo sapiens HRIHFB2216 mRNA, partial cds., N = 1, Score = 570, P = 2.9e-55

>TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds.
Length = 379

HSPs:

Score = 1202 (180.3 bits), Expect = 3.1e-122, P = 3.1e-122
Identities = 258/377 (68%), Positives = 283/377 (75%)

```

Query:   62 SPPSEEGGVQDPAGR-----GGTPQIRVVGGRGHVAIKAGQEE--GQP-P--AEGLAA 110
          SP +EG D G GTP R + G G+ G P P EGL
Sbjct:   3 SPERDEGTPVPDSRGHCDADTVSGTPDRRPLLGEKAVTGEGRAGIVGSPAPRDVEGLVP 62

Query:  111 ASVVMAADRSLLK-KVQGGEKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAE 169
          V AA + V+G A+ + ++ T GAE++A +VKT + TV+AA
Sbjct:  63 QIRVAAARQGESPPSVRGPAAVFVTPKYVEKAQETRGAESQARDVKT-EPGTVAAAA-- 119

Query:  170 RESAEVVVKEGLAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWLHEALR 229
          E +EV EE MEVE Q P GEE+E+ E EA EE GPW L LR
Sbjct:  120 -EKSEVATPGS-----EEVMEVE-QKPAGEEMEMLEASGGVREAPPEEAGPWLHGIDLR 170

Query:  230 MDPLEAIQLELDTVNAQADRAFFQLEHKFGRMRRLHYLERRNYIIQNIPGFWMATAFRNHPQ 289
          +PLEAIQLELDTVNAQADRAFFQ LE KFGMRRLHYLERRNYIIQNIPGFWMATAFRNHPQ
Sbjct:  171 RNPLEAIQLELDTVNAQADRAFFQLEHKFGRMRRLHYLERRNYIIQNIPGFWMATAFRNHPQ 230

Query:  290 LSAMIRGQDAEMLRITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLVKEYEVRSSGRV 349
          LSAMIRG+DAEMLRITNLEVKELRHP+TGCKFKFFFRNPNPYFRNKLVKEYEVRSSGRV
Sbjct:  231 LSAMIRGRDAEMLRITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLVKEYEVRSSGRV 290

Query:  350 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSHSLPESDKIAEIIKEDLWPNPLQYY 409
          VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSHSLPESD+IAEIIKEDLWPNPLQYY
Sbjct:  291 VSLSTPIIWRRGHEPQSFIRRNQDLICSFFTWFSHSLPESDRIAEIIKEDLWPNPLQYY 350

Query:  410 LLREGVRRARRRPLREPVEIPRPFQFQSG 438
          L REG+RR RRRP+REPVEIPRPFQFQSG
Sbjct:  351 LCREGIRRRRRRPIREPVEIPRPFQFQSG 379

```

Pedant information for DKFZphfbr2_2d15, frame 3

Report for DKFZphfbr2_2d15.3

```

[LENGTH]      438
[MW]           49307.65
[pI]           5.36
[HOMOL]        TREMBL:AF042180_1 gene: "Tspyl1"; product: "testis-specific Y-encoded-like
protein"; Mus musculus testis-specific Y-encoded-like protein (Tspyl1) mRNA, complete cds. 1e-
107
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR048c]
1e-07
[FUNCAT]       09.13 biogenesis of chromosome structure [S. cerevisiae, YKR048c] 1e-07
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR048c] 1e-07
[BLOCKS]       BL00376F
[PIRKW]        nucleus 6e-39
[PIRKW]        DNA binding 3e-06
[PIRKW]        phosphoprotein 6e-39
[PIRKW]        alternative splicing 6e-39
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY 22.83 %

```

```

SEQ      MSGLDGVKRTTPLQTHSIIISDQVPSQDAHQYLRLRDQSEATQVMAEPGEGGSETVALP
SEG      .....X
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      PSPPSEEGGVQDPAGRGGTPQIRVVGGRGHVAIKAGQEEGQPPAEGLAAASVVMAADRS
SEG      xxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LKKGVGQGEKALEICGAQRSASELTAGAEAEAEVKTGKCATVSAVAERESAEEVVVKEG
SEG      .....xxxxxxx.....xxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LAEKEVMEEQMEVEEQPPEGEEIEVAEEDRLEEEAREEEGWPWLHEALRMDPLEAIQLEL
SEG      .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      DTVNAQADRAFFQLEHKFGRMRRLHYLERRNYIIQNIPGFWMATAFRNHPQLSAMIRGQDAE
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      MLRYITNLEVKELRHPRTGCKFKFFFRNPNPYFRNKLVKEYEVRSSGRVSLSTPIIWR

```



```
SEG .....
PRD hhhhhhhhhhhhhccccceeeeeccccccchhhhhccccccccccccceeecc

SEQ GHEPQSFIRRNQDLICSFFTWFSDSLPESDKIAEIIKEDLWPNPLQYYLLREGVRRARR
SEG .....xxxxxxxxxxxx
PRD cccccchhhhhccccceeeeeccccccchhhhhhhccccceeeccccchhhh

SEQ RPLREPVEIPRPFGFQSG
SEG xxxxxxxx.....
PRD hcccccccccccccccc
```

(No Prosite data available for DKFZphfbr2_2d15.3)

(No Pfam data available for DKFZphfbr2_2d15.3)

DKFZphfbr2_2d17

group: transmembrane proteins

DKFZphfbr2_2d17 encodes a novel 292 amino acid protein with similarity to a C.elegans hypothetical protein.

One transmembrane region is predicted for the protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans hypothetical protein

TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 1009 bp

Poly A stretch at pos. 990, polyadenylation signal at pos. 969

```

1 TGGGCCTGTG GCTGGGGGCA GAGCTCAGAC TGTCTTCTGA AGATTGATGT
51 CTATTTCCCTT GAGCTCTTTA ATTTTGTTC CAATTTGGAT AAACATGGCA
101 CAAATCCAGC AGGGAGGTCC AGATGAAAAA GAAAAGACTA CCGCACTGAA
151 AGATTTATTA TCTAGGATAG ATTTGGATGA ACTAATGAAA AAAGATGAAC
201 CGCCTCTTGA TTTTCTGTAT ACCCTGGAAG GATTGAATA TGCTTTTAAT
251 GAAAAGGGAC AGTTAAGACA CATAAAAACT GGGGAACCAT TTGTTTTTAA
301 CTACCCGGAA GATTACACA GATGGAACCA GAAAAGATAC GAGGCTCTAG
351 GAGAGATCAT CACGAAGTAT GTATATGAGC TCCTGGAAAA GGATTGTAAT
401 TTGAAAAAAG TATCTATTCC AGTAGATGCC ACTGAGAGTG AACCAAGAG
451 TTTTATCTTT ATGAGTGAGG ATGCTTTGAC AAATCCACAG AAAGTATGG
501 TTTTAATTCA TGGTAGTGGT GTTGTCAAGG CAGGCGAGTG GGCTAGAAGA
551 CTTATTATTA ATGAAGATCT GGACAGTGGC ACACAGATAC CGTTTATTAA
601 AAGAGCTGTG GCTGAAGGAT ATGGAGTAAT AGTACTAAAT CCCAATGAAA
651 ACTATATTGA AGTAGAAAAG CCGAAGATAC ACGTACAGTC ATCATCTGAT
701 AGTTCAGATG AACCAGCAGA AAAACGGGAA AGAAAAGATA AAGTTTCTAA
751 AGTAACAAAG AAGCGACGTG ATTTCTATGA GAAGTATCGT AACCCCAAAA
801 GAGAAAAAGA AATGATGCAA TTGTATATCA GAGTGAGTGA GATCACTACT
851 TTCCTTTACT ATTTCTTTA CCTGTATAT ATTTATTAT ATGTAGATTG
901 TTTTGTTTTT CTCAAGAAT ATTAATTTCT TTATTTGTCA TCATTTATTT
951 CCCATGGTCG TCTACTTGA TTAATGGGT TTTAAATTC AAAAAAAAAA
1001 AAAAAAAAAA

```

BLAST Results

Entry I89937 from database EMBL:

Sequence 11 from patent US 5723315.

Score = 1083, P = 2.2e-42, identities = 223/231

Entry I89938 from database EMBL:

Sequence 12 from patent US 5723315.

Score = 875, P = 7.4e-33, identities = 175/175

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 47 bp to 922 bp; peptide length: 292

Category: similarity to unknown protein

Classification: unset

1 MSISLSSLIL LPIWINMAQI QQGGPDEKEK TTALKDLLSR IDLDELMKKD

```

51 EPPLDFPDTL EGFYAFNEK GQLRHIKTGE PFVFNRYEDL HRWNQKRYEA
101 LGEIITKYVY ELLEKDCNLK KVSIPVDATE SEPFSFIFMS EDALTNPQKL
151 MVLHSGGVV RAGQWARRLI INEDLDSTGQ IPFIKRAVAE GYGVIVLNP
201 ENYIEVEKPK IHVQSSSDSS DEPAEKREK DKVSKVTKKR RDFYEKYRNP
251 QREKEMMLY IRVSEITTFLL YFLYLVYIL LYVDCFVFLQ EY

```

BLASTP hits

Entry S67436 from database PIR:
 hypothetical protein - fission yeast (Schizosaccharomyces pombe)
 Length = 266
 Score = 112 (39.4 bits), Expect = 0.00037, P = 0.00037
 Identities = 33/147 (22%), Positives = 69/147 (46%)

Entry CEY75B8A_12 from database TREMBLNEW:
 gene: "Y75B8A.31"; Caenorhabditis elegans cosmid Y75B8A
 Score = 327, P = 1.5e-29, identities = 72/140, positives = 93/140

Alert BLASTP hits for DKFZphfbr2_2d17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2d17, frame 2

Report for DKFZphfbr2_2d17.2

```

[LENGTH]      292
[MW]           34260.50
[pI]           5.50
[HOMOL]        TREMBLNEW:AF064782_1 product: "unknown"; Mus musculus clone pEN87 unknown mRNA,
partial cds. 1e-119
[KW]           SIGNAL_PEPTIDE 19
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 10.96 %

```

```

SEQ  MSISLSSLILLPIWINMAQIQGGPDEKEKTTALKDLSRIDLDELMKKDEPPLDFPDTL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  EGFYAFNEKGQLRHIKTGEFVFNRYEDLHRWNQKRYEALGEIITKYVYELLEKDCNLK
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  KVSIPVDATESEPKSFIFMSDALTNPQKLMVLHSGGVVRAGQWARRLIINEDLDSTGQ
SEG  .....
PRD  eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
MEM  .....

SEQ  IPFIKRAVAEGYGVIVLNPENYIEVEKPKIHVQSSSDSSDEPAEKREKDKVSKVTKKR
SEG  .....
PRD  chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....

SEQ  RDFYEKYRNPQREKEMMLYIRVSEITTFLLYFLYLVYILLYVDCFVFLQY
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

(No Prosite data available for DKFZphfbr2_2d17.2)

(No Pfam data available for DKFZphfbr2_2d17.2)

DKFZphfbr2_2d20

group: brain derived

DKFZphfbr2_2d20 encodes a novel 197 amino acid protein with similarity to *Synechocystis* sp. P74594 hypothetical32.8 kD protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *Synechocystis* sp. (PCC 6803)

complete cDNA, complete cds, EST hits
potential start at bp 67 matches kozak consensus ANCatgG

Sequenced by Qiagen

Locus: unknown

Insert length: 1787 bp

Poly A stretch at pos. 1768, polyadenylation signal at pos. 1743

```
1 TGGGGCGGCC GCGGCGGGAA CATGGAGGAG CTGCTGAGGC GCGAGCTGGG
51 CTGCAGCTCT GTCAAGGCCA CGGGCCACTC GGGGGGCGGG TGCATCAGCC
101 AGGGCCGGAG CTACGACACG GATCAAGGAC GAGTGTTCGT GAAAGTGAAC
151 CCCAAGCCGG AGGCCAGAAG AATGTTTGAA GGTGAGATGG CAAGTTTAAC
201 TGCCATCCTG AAAACAAACA CGGTGAAAGT GCCCAAGCCC ATCAAGGTTT
251 TGGATGCCCC AGGCGGCGGG AGCGTGCTGG TGATGGAGCA CATGGACATG
301 AGGCATCTGA GCAGTCATGC TGCAAAGCTT GGAGCCGAGC TGGCCGATT
351 ACACCTTGAT AACAGAAGC TTGGAGAGAT GCGCCTGAAG GAGGCGGGCA
401 CAGTGTGGAG AGGAGGTGGG CAGGAGGAAC GGCCTTTGT GGCCTGGTTT
451 GGATTTGACG TGGTGACGTG CTGTGGATAC CTCCCCAGG TGAATGACTG
501 GCAGGAGGAC TGGGTCGTGT TCTATGCCCG GCAGCGCATT CAGCCCCAGA
551 TGGACATGGT GGAGAAGGAG TCTGGGGACA GGGAGGCCCT CCAGCTTTGG
601 TCTGCTCTGC AGTAAAGAT CCCTGACCTG TTCCGTGACC TGGAGATCAT
651 CCCAGCCTTA CTCCACGGGG ACCTCTGGGG TGGAAACGTA GCAGAGGATT
701 CCTCTGGGCC GGTGATTTT GACCCAGCTT CTTTCTACGG CCACTCGGAA
751 TATGAGCTGG CAATAGCTGG CATGTTTGGG GGCTTTAGCA GCTCCTTTTA
801 CTCCGCCTAC CACGGCAAAA TCCCCAAGGC CCCAGGATTC GAGAAGCGCC
851 TTCAGTTGTA TCAGCTCTTT CACTACTTGA ACCACTGGAA TCATTTTGGA
901 TCGGGGTACA GAGGATCCTC CCTGAACATC ATGAGGAATC TGGTCAAGTG
951 AGCGGGCCTT ACTCTGGAAG GAGGTCTCAG AGGTTTCTCC ACAGTCCTCT
1001 TCTGGGCAAA TTCTTGTTTC TTCACATGCC GGACTAGCTT AAGACCAATG
1051 CAGTAGCTTA TTCCAAGCC TTGCAAAGTA TATAATATCT AAGAGGAAAG
1101 GTTTTGTCTAT CCCAGCGTTG TCCACTTTGT GGGGCTTTGT AGGTAGACGG
1151 AGCCACACTA CAGGCAGGGT ATGAGCAGAG GGATGTATGG AGTGTGGGCG
1201 ACTCTGAGCC TACTGCTGTC TGCAAGGTGG GGAACTGTA AGTGAACCCC
1251 TGTGGGTGCG GGGGAGGGTA TCCGGTGCGC AGGGAGGTGG CCAGCGCCCC
1301 CGGGCACTGC TGCTCATAGG TACCTTTCCG CTGCCTCCTC CCGTCTCTCC
1351 TGTGAGGAA TGCTCTGAG CTGTTACAGT TGATGCTTCT TGGTTGGCAA
1401 GACTTGGGTG TAGACATGAA ACCACCTTAC TAAAAGCGTC TAAAATGAC
1451 CAATTCCAGA ATCAAGCGTA TTCCGTTTTC CTCCTGCATG ATCCCTGGGC
1501 CCTCCGCGAG GCTGAGCAAG TCTGTAACT GATTCTGGGA GAAACCAAGC
1551 TGCTGGCCGT AGGATGTCCT TGGGTACATC CAGGAGTCTT CATTGCTTCT
1601 GTTATTACCC CGTCTCCTCT GCCATTTTCT ACAGCTTGCT GAGTTGTCAT
1651 TCCTTTGCAA CATTAAATA CATGCTGAC TCATATTTT CCTTCCTTCA
1701 CTGTTGTAGT AAAGAGACAT ATTTATGAA TGGCATTGAT GCTAATAAAC
1751 CCTTTGCCCA AAAATTGAA AAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

1 MEELLRRELG CSSVRATGHS GGGCISQGRS YTDQGRFV KVPKAEARR
51 MFEGEAMSL AILKNTTKV PKPIKVLDP GGGSVLMEH MDMRHLSSHA
101 AKLGAQSLADL HLDNKKLGEM RLKEACTVWR KGGQEERPF ARFGEDVUTC
151 CRYLPQVNDW QEDWVVFYAR QRIPQMDMV EKGSGDREAL QLWSALQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2d20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2 2d20, frame 1

Report for DKFZphfbr2 2d20.1

```
[LENGTH]      197
[MW]           21963.25
[pI]           6.96
[HOMOL]        PTR:S76790 hypothetical protein - Synechocystis sp. (strain PCC 6803) 9e-12
```

```
[SUPFAM]      hypothetical protein b1725 1e-06
[PROSITE]     LEUCINE_ZIPPER 1
[PROSITE]     MYRISTYL 2
[PROSITE]     GLYCOSAMINOGLYCAN 1
[PROSITE]     PKC_PHOSPHO_SITE 2
[KW]          Alpha Beta
```

SEQ MEELLRRELGCSSVRATGHSGGGCISQGRSYDTDQGRVFKVNPKEARRMFEGEMASLT
PRD ccchhhhhcccccceeecccccceeecccccceeeccchhhhhhhhhhhhhhh

SEQ AILKTNVTKVPKPIKVLDAPGGGSVLVMEHMDMRHLSSHAAKLGAQLADLHLDNKKLGEM
PRD hhhhhhheeeccceeecccccccceeecccccccchhhhhhhhhhhhhhhccccchhh

SEQ RLKEAGTVWRGGGQEERPFVARFGFDVVTCCGYLPQVNDWQEDWVVFYARQRIQPQMDMV
PRD hhhhhccccccccccccceeeccccceeeccccccccccccchhhhhhhhhhhhhhhhhhh

```

SEQ      EKESGDREALQLWSALQ
PRD      hhhccchhhhhhhhhccc

```

Prosites for DKFZphfbr2_2d20.1

PS00002	20->24	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	13->16	PKC_PHOSPHO_SITE	PDOC00005
PS00005	67->70	PKC_PHOSPHO_SITE	PDOC00005
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00029	96->118	LEUCINE ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_2d20.1)

DKFZphfbr2_2g18

group: brain derived

DKFZphfbr2_2g18 encodes a novel 229 amino acid protein with partial similarity to the humane dJ30M3.2 gene product.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

J30M3.2 extension of genmodel

complete cDNA, complete cds, EST hits
(mouse ESTs with >90% Identities)

Sequenced by Qiagen

Locus: /map="6p22.1-22"

Insert length: 2444 bp

Poly A stretch at pos. 2425, no polyadenylation signal found

```
1 TGGTCGAGGG TCGACGGTAT CGATAAGTTT TTTTTTTTTT TTTTTTTTTT
51 TGGAAAGCAA GGATCACACT TCCCCCTCCC TGTTCCCTAA TCCCTTTTCT
101 AAAAAAGGGG GAAATCCCGG ATGGATTTTA GGGATTGGTC TGGTGTGAGC
151 TGTGTCTTAT TGCACACCTA AATCCTGATT ATAGGCTTTT CATTCTCCCG
201 CAAAGCCTTT ATTTTGGCAG TTAAGCCAAA TGTGTTTTC AGAAAGTTAG
251 TTATTTTCTC CTCTTCTTTT CCTTCTTTTC CTCCCTTTT CCCGCTGAC
301 CCCAAACGTT ATTGTCCAAA CATGACTGGA CAGCAGCTTT TGTTTCTTGA
351 CCCTGTAATA TGACAGTCTG CTAATATTGA CAGAAGGTGC AGTTTTTGGG
401 TTATAGTCTG GATTTTCGCT AATCAATCAT ATTAGCAGGA AAAAAAATGA
451 CTTGTTTCTG TTGTACTTGA GTCTTAAGAA AAAGTGCCCA TAGTTTAGTG
501 ACAATTTCCA AAGGCTTTAG TACCACCTGT ATTTCAAAT GGGGGACCCA
551 AACTCCCGGA AGAAACAAGC TCTGAACAGA CTACGTGCTC AGCTTAGAAA
601 GAAAAAAGAA TCTCTAGCTG ACCAGTTTGA CTTCAAGATG TATATTGCCT
651 TTGTATTCAA GGAGAAGAAG AAAAAGTCAG CACTTTTGA AGTGTCTGAG
701 GTTATACCAG TCATGACAAA TAATTATGAA GAAAATATCC TGAAAGGTGT
751 GCGAGATTCC AGCTATTCCT TGGAAAGTTC CCTAGAGCTT TTACAGAAGG
801 ATGTGGTACA GCTCCATGCT CCTCGATATC AGTCTATGAG AAGGGATGTA
851 ATTGGCTGTA CTCAGGAGAT GGATTTTCATT CTTTGGCCTC GGAATGATAT
901 TGAAAAAATC GTCTGTCTCC TGTTTTCTAG GTGGAAAGAA TCTGATGAGC
951 CTTTTAGGCC TGTTCCAGCC AAATTTGAGT TTCATCATGG TGAATATGAA
1001 AAACAGTTTC TGCATGTACT GAGCCGCAAG GACAAGACTG GAATCGTTGT
1051 CAACAATCCT AACCACTCAG TGTTTCTCTT CATTGACAGA CAGCACTTGC
1101 AGACTCCAAA AAACAAAGCT ACAATCTTCA AGTTATGCAG CATCTGCCTC
1151 TACCTGCCAC AGGAACAGCT CACCCACTGG GCAGTTGGCA CCATAGAGGA
1201 TCACCTCCGT CCTTATATGC CAGAGTAGAG TACTGACCAG CAAATGGAG
1251 AAGATCAGAG AATGCAGCAG CAGTTTTTTT TCTTGTTTTC TTACCACTTT
1301 ATCTTTTCAG AGTTTAAAGA AAATGGACTC ATGCACAGAA CACTATGCAT
1351 TTTGAAACTT GTTCATCTCG GATTTTTTTA AATCATTTTT ATCTCAGAAC
1401 TTAACAAAAA ATTAGATGTC GTGCACGGAC TGTGTGAAAG AAGATGCTTT
1451 GCATATTTGC TGCATGTCAT CAGTATCTTA CTAATAATGT GAAATGAAAG
1501 GACTATTGTA CACTGAAATG CTTAAATGTA TCTGAAAGCA CAAGGTGATA
1551 CTCATTTTTA TGGTCTTCCC ATTTGTGCTG GTTTTTGCCT CTTTGACATC
1601 TGTCATCAGT ATTTAGAGGG TGAGAAAGTA ATGTAACAGG TATAAATAAC
1651 ATTTTAAAAA ACAATAACTT TGCTATAATC ACAGTTGTTC CAGAGCACTG
1701 TCAGATACAT TCTAATGACC AGAACTGGTT TAAAAAAGA AAATACAACC
1751 ATGGGAAAGA AATCTTAAAT GAAAAACGCA TCTCATTGTA GGCATTTTTG
1801 CCTCATATTT TACTGGGCCA TGTTTGTTTC CTGGTACTCA TGTATTTTTT
1851 TTTTTCCTAG ATCTCTTTCC CCAAGTTGCT ATTGTAAGAG TATCTGCTG
1901 CGTGTGGATG CAGTTATACA CATTAAAGCA GATCTGGAGT CTGAAGTAGC
1951 TATAAAGCAG CTATAAACA GAAATACATG CATAGCTGCA GAAACCATGA
2001 TAGGTAGAGG ACTTTTCTTT TGGTTTTGTT TTGTTTTGTT TTGTTTTGTT
2051 TTTGGTTTTA CAGAGAAGAG ATTTTATTA CAAAGAAAAA AATCCAGTG
2101 AATTGTGCAG AAATGCTGGT TTTTACACCA TCCTAAAGAA AAACTTTACA
2151 AGGGTGTGTT GGAGTAGAAA AAAGGTTATA AAGTTGGAAT CTTAAATTGT
2201 AAAATTAAAC ATTGAGTGT CAAAGTTCTAA AAGCAGAACT CATTTCGTGC
2251 AATGAACATA AGGAAAGACT ACTGTATAGG TTTTTTTTTT TCTCCTTTTA
2301 AATGAAGAAA AGCTTTGCTT AAGGGTTGCA TACTTTTATT GGAGTAAATC
2351 TGAATGATCC TACTCCTTG GAGTAAGACT AGTGCTTACC AGTTTCCAAT
2401 TGTATTTAGC TTCTGTTGGA ATTTGAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS338352 from database EMBL:
human STS EST171398.
Score = 1747, P = 3.0e-74, identities = 359/365

Entry HS447255 from database EMBL:
human STS SHGC-10143.
Score = 1717, P = 6.5e-73, identities = 365/383

Entry HS30M3 from database EMBLNEW:
Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Score = 6646, P = 0.0e+00, identities = 1344/1355

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 539 bp to 1225 bp; peptide length: 229
Category: putative protein

```

1 MGDPNRKKQ ALNRLRAQLR KKKESLADQF DFKMYIAFVF KEKKKKSALF
51 EVSEVIPVMT NNYEENILKG VRDSSYSLES SLELLQKDVV QLHAPRYQSM
101 RRDVIGCTQE MDFILWPRND IEKIVCLLFS RWKESDEPFR PVQAKFEFHH
151 GDYEKQFLHV LSRKDKTGIV VNNPNQSVFL FIDRQHLQTP KNKATIFKLC
201 SICLYLPQEQ LTHWAVGTIE DHLRPYMPE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2g18, frame 2

TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands., N = 1, Score = 470, P = 1.1e-44

>TREMBLNEW:HS30M3_2 gene: "dJ30M3.2"; product: "dJ30M3.2 (novel protein)"; Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3. Contains three novel genes, one similar to C. elegans Y63D3A.4 and one similar to (predicted) plant, worm, yeast and archaea bacterial genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSSs and putative CpG islands.
Length = 86

HSPs:

Score = 470 (70.5 bits), Expect = 1.1e-44, P = 1.1e-44
Identities = 86/86 (100%), Positives = 86/86 (100%)

```

Query: 144 AKFEFHGDDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 203
        AKFEFHGDDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC
Sbjct: 1 AKFEFHGDDYEKQFLHVLRSRKDKTGIVVNNPNQSVFLFIDRQHLQTPKNKATIFKLCSIC 60

Query: 204 LYLPQEQ LTHWAVGTIEDHLRPMPE 229
        LYLPQEQ LTHWAVGTIEDHLRPMPE
Sbjct: 61 LYLPQEQ LTHWAVGTIEDHLRPMPE 86

```

Pedant information for DKFZphfbr2_2g18, frame 2

Report for DKFZphfbr2_2g18.2

Prosites for DKFZphfbr2_2g18.2

(No Pfam data available for DKFZphfbr2_2g18.2)

DKFZphfbr2_2h1

group: brain derived

DKFZphfbr2_2h1 encodes a novel 180 amino acid protein with weak similarity to C.elegans D2007.4 protein

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C.elegans D2007.4 protein

CpG island in 5' region, complete cDNA

Sequenced by Qiagen

Locus: unknown

Insert length: 957 bp

Poly A stretch at pos. 939, polyadenylation signal at pos. 916

```

1  GGGGGTCCCT GACTTTATAT GGCTGCTCCT GGCGAGCGAC TGAGTCGTCC
51  GTGAGGAAAA AGAGGCGAGG CTTTCCGAG ATCGTCTCAG CGATGGCGCT
101 TCGGTCGCGG TTTTGGGGT TGTCTCGGT TTGCAGGAAC CCTGGGTGCA
151 GGTTGCGAGC CCTGTCAACC AGCTCCGAGC CGGCAGCGAA ACCTGAAAGTG
201 GACCCCTGTG AAAATGAAGC TGTGCGCCCA GAATTCACCA ACCGGAACCC
251 CCGGAACCTG GAGCTTTTGT CTGTAGCCAG GAAAGAGCGG GGCTGGCGGA
301 CGGTGTTTCC CTCCCGTGAG TTCTGGCACA GGTGCGGAGT TATAAGGACT
351 CAGCATCATG TAGAAGCACT TGTGGAGCAT CAGAATGGCA AGGTTGTGGT
401 TTCGGCCTCC ACTCGTGAGT GGGCTATTAA AAAGCACCTT TATAGTACCA
451 GAAATGTGGT GGCTTGTGAG AGTATAGGAC GAGTGCTGGC ACAGAGATGC
501 TTAGAGGCGG GAATCAACTT CATGGTCTAC CAACCAACCC CGTGGGAGGC
551 AGCCTCAGAC TCGATGAAAC GACTACAAAG TGCCATGACA GAAGGTGGTG
601 TGGTTCTACG GGAACCTCAG AGAATCTATG AATAAATGGA AGCATTAATT
651 GTTTTGAACA TGTAAATATA AATCTGTGAG CCACTACAGC CATCAAAAGA
701 GAGCATCTGG AAGAACAGCC AGCTTGGAAG TTTTACAGCA ATAATGTTGC
751 AGTGGAATAT TATTTGTAGT TAAGGTCATC CTCCTCCCCT TTCTGTTTTT
801 TTAATCAAG AACTACGTTT TGCCCTCTCT TTGGGCTTCA GAAGCATCTA
851 AGAAAAGCAG TCATCAATTA TAATTAACCT TCAAAGGGCA AGTCAGAAAGT
901 TGTTTATAAA TTACAAAATA AAGGCATATT ATGAACCTTA AAAAAAAAAA
951 AAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 93 bp to 632 bp; peptide length: 180
 Category: similarity to known protein
 Classification: unset

```

1  MALRSRFWGL FSVCRNPGCR FAALSTSSEP AAKPEVDPVE NEAVAPEFTN
51  RNPRLLELLS VARKERGWRT VFPSREFWHR LRVIRTQHHV EALVEHQNGK
101 VVVSASTREW AIKKHLYSTR NVVACESIGR VLAQRCLEAG INFVYQPTP
151 WEAASDSMKR LQSAMTEGGV VLREPQRIYE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2h1, frame 3

PIR:S44789 D2007.4 protein - *Caenorhabditis elegans*, N = 1, Score = 194, P = 2e-15

PIR:JC5753 ribosomal protein L18 - *Vibrio proteolyticus*, N = 1, Score = 121, P = 1.1e-07

>PIR:S44789 D2007.4 protein - *Caenorhabditis elegans*
Length = 170

HSPs:

Score = 194 (29.1 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 51/134 (38%), Positives = 78/134 (58%)

Query: 48 FTNRNPRNLELLSVARKERGWRVFP--SREFWHRLRVIRTQHHVEA-LVEHQNGKVVVS 104
F NRNPRN EL+ G++ +R + +++ ++ + H E LV +Q+G VV+S
Sbjct: 9 FVNRNPRNLMGRQAPNTGYQFEKDRAARSYIYKVELVEGKSHREGRLVHYQDG-VVIS 67

Query: 105 ASTREWAIKKHLYSTRNVVACESIGRVLAQRCLEAGINFVYQPTPWEAASDSMKRLQ-- 162
AST+E +I LYS + A +IGRVLA RCL++GI+F + T EA S +
Sbjct: 68 ASTKEPSIASQLYSKTDTSAALNIGRVLALRCLQSGIHFAMPGATK-EAIEKSQHQTHFF 126

Query: 163 SAMTEGGVVLREPQRI 178
A+ E G+ L+EP +
Sbjct: 127 KALEEEGLTLKEPAHV 142

Pedant information for DKFZphfbr2_2h1, frame 3

Report for DKFZphfbr2_2h1.3

[LENGTH] 180
[MW] 20576.57
[pI] 9.63
[HOMOL] PIR:S44789 D2007.4 protein - *Caenorhabditis elegans* 2e-13
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0794] 2e-04
[SUPFAM] Escherichia coli ribosomal protein L18 8e-06
[KW] Alpha_Beta

SEQ MALRSRFWGLFSVCRNPGCRFAALSTSSEPAAKPEVDPVENEVAPEFTNRNPRNLELLS
PRD cccccceeeeeccccceccccccccccccccccccccccccccccccccchhhh

SEQ VARKERGWRVFPFSREFWHRLRVIRTQHHVEALVEHQNGKVVVSASTREWAIKKHLYSTR
PRD hhhccccccccchhhhhhhhhccccchhhhhhhhhccccceeechhhhhhhhhhhcc

SEQ NVVACESIGRVLAQRCLEAGINFVYQPTPWEAASDSMKRLQSAMTEGGVVLREPQRIYE
PRD ccceeehhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhcccccecccccccc

(No Prosite data available for DKFZphfbr2_2h1.3)

(No Pfam data available for DKFZphfbr2_2h1.3)

DKFZphfbr2_2h10

group: brain derived

DKFZphfbr2_2h10 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 2176 bp

Poly A stretch at pos. 2161, polyadenylation signal at pos. 2143

```
1 TGGGGAGTAT TCTAATTATA TTTTATATTT AATAAATTAT TTTTCTATTT
51 CTTTGTGTTATA TTAAGTTGCA CACTTGTTTC TTTTATCCAG AAAAGTTAGT
101 ATAATAAAAA TAGTTTTAAG ATTAACGTGT AATGTAAAGG AAAAGTATTA
151 TTAATTATTT CAGGAAATTG CAAGACCTAA CATGGCTGAA AGAGAAACAG
201 AAACATCAAA TTCTGAAAGT AAACAAGATA AAGCTGCTTC TTCAAAAGAA
251 AAAAATGGAT GTAATGCAAA TTCATTGAA GGCTCATCAA CAACAAAAG
301 TGAAGAAAGC ATAACAGTTT CAGATAAGGA AAATGAAACC TGTCTTGCG
351 ACCAGGAAAC TGGCTCAAAA AACATCGTCA GTTGTGATTC AAATATTTGGT
401 GCAGATAAAG TGGAAAAGAA AAAACAAATA CAACACGTTT GTCAGGAAAT
451 GGAGTTGAAG ATGTGCCAGA GTTCAGAAAA CATAATCTTA TCTGATCAGA
501 TTAAGATCA CAACTCCAGT GAAGCCAGAT TTTCTTCAAA GAATATTAAG
551 GATTTGCGAT TAGCATCAGA TAATGTAAGC ATTGATCAGT TTTTGAGAAA
601 AAGACATGAA CCTGAATCTG TTAGTTCTGA TGTTAGCGAG CAAGGCAGTA
651 TTCATTGGA ACCTCTGACT CCATCCGAGG TACTTGAGTA TGAAGCCACA
701 GAGATTCTTC AGAAAGGTAG TGGTGATCCT TCAGCCAAGA CTGATGAAGT
751 AGTGTCTGAT CAAACAGATG ACATTCTCGG AGGAAATAAC CCTAGCACAA
801 CAGAGGCCAAC AGTAGACCTG GAAGATGAAA AAGAAAGAAG TTGAAATTAG
851 TCATTTTAAG TTTCAAGTGA CCAACGATAA GGCATTTTGG AACAGTGCTA
901 TCAGGTGAGC TCAGTGGTGC TGTGTGATGT TCAGAAATGG AAATATGTAA
951 GGGAGGTCAC ACATACACTT TACCTGTATG TTCAACCTAT GTTATCAAA
1001 AAACCAATTC ACCAATAATA GCATGATTAG TAGGGATTCC CAAAAAGTTT
1051 TTA AAAACAC GAACAGGATT TTAATGATAA TTA AATTGTC AGTGGAAAGG
1101 TCTCATTTAA TGGTTTTCAA GGAAATGGGA TTTGGTTGCT GACATGAATT
1151 GATGATATTA GTAATATTTA TAAAGCCTTT CAAACTTCCA TCAATCCTAA
1201 GCTAAAAATC TTTATTACCT GTATATCCTT TTCAGTTAAC TGAGAGGAAG
1251 GGATTTGGAA ACCATGTACT TTTGGGGAGT AATTGATTAA AAACAATGGC
1301 TGATTGGCAT TGTTAATGAA GGCTTTATTT GTGAGGATGA TGCTGGTAAA
1351 TGGAGCATGC TTAGAGTACT AAATTGATCT AATGAGAATT TGGATGAACA
1401 TAACTTTAAT TTTGGATTTA ATATAACATT CCAGTCAGAC GCATGTA AAC
1451 AGAATATTTG AATCTTTGTA CCTCCATACA AGTGTAGGCC TGCCAGGCTG
1501 TAAGCTTACC TTAATTA AAC TTTCAAGTGAA AGTGGAAATTA TTAAGATATA
1551 AATTATATTT TGTGCTTTTT GTCAGTGTGT AAGCTGTGTA GAAATTCCTT
1601 GATGTATTAG TTGTATTAAT GTAAAGTAGA AACCATTGTG TGAACTCCT
1651 GTAGCTATTA TGCTTTTAAT ATTTGTTTAA TGTTCTTCCT TAGAAATAGG
1701 CCCATAAAAA TGGTCTGGAA GCCAAACCAA AGTATGGTAT AATGTAGATA
1751 TTGTAAAGCA GTAACTGAA AACATGTCCT GGCATGTATT CAGCCATGTT
1801 TAAGTGACTT TTCTGTAATT GTAAATAAAA AACTTCAAAT GGGACCTAAA
1851 ACAGTGATGT AAAAGAACTG GTTTTGGAAA TTTAGCCTAA TTTATCTATA
1901 AGATGGCTGC TAAATTGATT TTTCAAGTTCT TTTTATCATC TAAATATAA
1951 TAGATATAGA AATGAATAAT ATGAAGAACA GTAGTTTGCT TTGAAATACT
2001 AATAAACTTT TATTTAAGAT GCTTCATTTT TACTTCTTAA AACGTGCTTT
2051 GGATTCCTAA ATTTTGTTC ACTGAATGTT CAATGTTTTA AATGGCGATT
2101 AAAATACTCT GCTGTATATA GTAGTTTTTG AGTAAATAT TGCATAAAAA
2151 ATCTGCCCCC GAAAAA AAAA
```

BLAST Results

Entry G35287 from database EMBL:

human STS SHGC-37375.

Score = 2163, P = 2.8e-91, identities = 437/441

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 182 bp to 841 bp; peptide length: 220
Category: putative protein

```

1 MAERETETSN SESKQDKAAS SKEKNGCNAN SFEGSSTTKS EESITVSDKE
51 NETCLADQED GSKNIVSCDS NIGADVEKK KQIQHVCQEM ELKMCQSSSE
101 IILSDQIKDH NSSEARFSSK NIKDLRLASD NVSIDQFLRK RHEPESVSSD
151 VSEQSGSHLE PLTPSEVLEY EATEILQKGS GDPSAKTDEV VSDQTDIDFG
201 GNNPSTTEAT VDLEDEKERS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2h10, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_2h10, frame 2

Report for DKFZphfbr2_2h10.2

```
[LENGTH]      220
[MW]           24109.02
[pI]           4.51
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YKR092c] 4e-05
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKR092c] 4e-05
[PROSITE]      MYRISTYL 3
[PROSITE]      CK2_PHOSPHO_SITE 8
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           Alpha Beta
```

```
SEQ      MAERETETSNSESKQDKAASSKEKNGCNANSFEGSSTTKSEESITVSDKENETCLADQET
PRD      cccccccccccccchhhhhhhhhccccccccccccccccceeeeeeccccccccccccccc

SEQ      GSKNIVSCDSNIGADKVEKKQIQHVQCEMELKMCQSSENIILSDQIKDHNSSSEARFSSK
PRD      cccceeeecccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeccccccccccccccccccc

SEQ      NIKDLRLASDNVSIQDLFRKRHEPESVSSDVSEQSGSIHLEPLTPSEVLEYEATEILQKGS
PRD      cchhhhhhhccchhhhhhhhhccccccccccccccccceecccccccchhhhhhhcccccc

SEQ      GDPSAKTDEVVSDQTDIPGGNNPSTTEATVLDLEDEKERS
PRD      cccccccccccccccccceehhhhhhhhhcc
```

Prosite for DKFZphfbr2 2h10.2

PS000001	51->55	ASN_GLYCOSYLATION	PDOC000001
PS000001	111->115	ASN_GLYCOSYLATION	PDOC000001
PS000001	131->135	ASN_GLYCOSYLATION	PDOC000001
PS000005	20->23	PKC_PHOSPHO_SITE	PDOC000005
PS000005	37->40	PKC_PHOSPHO_SITE	PDOC000005
PS000005	47->50	PKC_PHOSPHO_SITE	PDOC000005
PS000005	118->121	PKC_PHOSPHO_SITE	PDOC000005
PS000005	184->187	PKC_PHOSPHO_SITE	PDOC000005
PS000006	9->13	CK2_PHOSPHO_SITE	PDOC000006
PS000006	13->17	CK2_PHOSPHO_SITE	PDOC000006
PS000006	20->24	CK2_PHOSPHO_SITE	PDOC000006
PS000006	38->42	CK2_PHOSPHO_SITE	PDOC000006
PS000006	45->49	CK2_PHOSPHO_SITE	PDOC000006
PS000006	47->51	CK2_PHOSPHO_SITE	PDOC000006
PS000006	163->167	CK2_PHOSPHO_SITE	PDOC000006
PS000006	205->209	CK2_PHOSPHO_SITE	PDOC000006
PS000008	26->32	MYRISTYL	PDOC000008

PS00008	34->40	MYRISTYL	PDOC00008
PS00008	201->207	MYRISTYL	PDOC00008

Pfam for DKFZphfbr2_2h10.2

HMM_NAME	TNFR/NGFR cysteine-rich region		
HMM	*CpeG.tYtD.WNHvpqClpCtrCePEMGQYMvgPCTwTQNTVC*		
	+E+ T +D +N ++C E G+ + +C+++ +		
Query	40	SEESITVSDKEN--ETC--LADQET--GSKNIVSCDSNIGADK	76

DKFZphfbr2_2il7

group: intracellular transport and trafficking

DKFZphfbr2_2il7.3 encodes a novel 201 amino acid putative GTP-binding protein related to Rab1B.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory(biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes. Rab1B is essential for the intracellular transport of nascent low density lipoprotein (LDL) receptor. It is discussed as a universal mediator of endoplasmatic reticulum to Golgi transport of membrane glycoproteins in mammalian cells.

The new protein can find clinical application in modulating the transport of glycoproteins inside cells, especially of the LDL receptor.

Medline

96245776: Intracellular transport and maturation of nascent low density lipoprotein receptor is blocked by mutation in the Ras-related GTP-binding protein, RAB1B

strong similarity to rab1

complete cDNA, complete cds, start at 47, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1985 bp

Poly A stretch at pos. 1901, polyadenylation signal at pos. 1859

```
1 GGGAGCAGAG TCGACTGGGA GCGACCGAGC GGGCCGCCGC CGCCGCCATG
51 AACCCCGAAT ATGACTACCT GTTTAAGCTG CTTTGTATG GCGACTCAGG
101 CGTGGGCAAG TCATGCCTGC TCCTGCGGTT TGCTGATGAC ACGTACACAG
151 AGAGCTACAT CAGCACCATC GGGGTGGACT TCAAGATCCG AACCATCGAG
201 CTGGATGGCA AACTATCAA ACTTCAGATC TGGGACACAG CGGGCCAGGA
251 ACGGTTCGGG ACCATCACTT CCAGCTACTA CCGGGGGGCT CATGGCATCA
301 TCGTGGTGTA TGACGTCACT GACCAGGAAT CCTACGCCAA CGTGAAGCAG
351 TGCTGTCAGG AGATTGACCG CTATGCCAGC GAGAACGTCA ATAAGCTCCT
401 GGTGGGCAAC AAGAGCGACC TCACCACCAA GAAGTGGTG GACAACACCA
451 CAGCCAAGGA GTTGCAGAC TCTCTGGGCA TCCCTTCTT GGAGACGAGC
501 GCCAAGAATG CCACCAATGT CGAGCAGGCG TTCATGACCA TGGCTGCTGA
551 AATCAAAAAG CGGATGGGGC CTGGAGCAGC CTCTGGGGGC GAGCGGCCCA
601 ATCTCAAGAT CGACAGCACC CCTGTAAAGC CGGCTGGCGG TGGCTGTTGC
651 TAGGAGGGGC ACATGGAGTG GGACAGGAGG GGGCACCTTC TCCAGATGAT
701 GTCCCTGGAG GGGGAGGAG GTACCTCCCT CTCCTCTCC TGGGGCATTT
751 GAGTCTGTGG CTTTGGGGTG TCCTGGGCTC CCCATCTCCT TCTGGCCCAT
801 CTGCCTGCTG CCCTGAGCCC CGGTTCTGTC AGGGTCCCTA AGGGAGGACA
851 CTCAGGGCCT GTGGCCAGGC AGGGCGGAGG CCTGCTGTGC AGTTGCCCTCT
901 AGGTGACTTT CCAAGATGCC CCCCTACACA CCTTCTTTG GAACGAGGGC
951 TCTTCTGTGC GTGTCCCTCC CACCCCATG TATGCTGCAC TGGGTTCTCT
1001 CCTTCTTCTT CTGTCTGTCC TGCCCAAGAA CTGAGGGTCT CCCCGGCCCTC
1051 TACTGCCCTG GCTGCAGTCA GTGCCAGGG CGAGGAATGT GGCCAGGGGA
1101 TCCAGGACCT GGGATCCAGG GCCCTGGGCT GGACCTCAGG ACAGGCATGG
1151 AGGCCACAGG GGCCAGCAG CCCACCCCTT CCTCTCCCA CTGCCTCCTC
1201 TCCCTTCCTA CACTCCCAGC TCGAGCCGTC CAGCTGCGGT GGGATCTGAG
1251 TATATCTAGG GCGGGTGGGC GGGTAGCAGT GCTGGGCCCTG TGTCTTGAGC
1301 CTGGAGGGAG ACTGCTCCTG CCGCCCTCTG CCTTGCCGGA GACAGACCCA
1351 TGGCGTGCCT GCCACCCGTG CCCCTTTGTC CCCATGTGAG GCGGAGGCGG
1401 AAGGCCACAC GTGCCAGAGG CTGGGCACCA GCCTTAACCC TCACTCTGCT
1451 AGCACCTCCT CCCTTTCCCC AAGGTAGCAC ATCTGGCTCA CTCCCCACTC
1501 CGTCTCTGGA GCCCACCAGG GAAGGCCCTC ATCCCCTGCC GCTACTTCTC
1551 TGGGGAATGT GGGTTCCATC CAGGATTGGG GGCCTCTCTG CTCACCCACT
1601 CTGCACCCAG GATCTAGTC CCCTGCCCTC TGGCACAGCT GCTTCTGCA
1651 AGAAAGCAAG TCTTTGGTCT CCCTGAGAAG CCATGTCCCT CGTGCTGTCT
1701 CTTGCTGTGC CCACCTGTGC CCTGCCCTCC AGCTTGTATT TAAGTCCCTG
1751 GGTGTGCCCC TTGGGGTGCC CCCCCTCCC AGGTTCCTCT CTGGTGTGAT
1801 GTCAGGCATT TTGCAAGGAA AAGCCACTTG GGGAAAGATG GAAAGGACA
1851 AAAAAAATTA ATAAATTCC ATTGGCCCTC GGGTGAGCTG AGGGTTTTTG
1901 CAAGGAAAAA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
1951 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
```

BLAST Results

No BLAST result

Medline entries

91115900:
A family of ras-like GTP-binding proteins expressed in electromotor neurons.

Peptide information for frame 3

ORF from 48 bp to 650 bp; peptide length: 201
Category: strong similarity to known protein

1 MNPEYDYLFK LLLIGDSGVG KSCLLLRFAD DTYTESYIST IGVDKIRTI
51 ELDGKTIKIQ IWDTAGQERF RTITSSYYRG AHGIIVVYDV TDQESYANVK
101 QWLQEIERYA SENVNKLLVG NKSDLTTKKV VDNTAKEFA DSLGIPFLET
151 SAKNATNVEQ AFMTMAAEIK KRMGPGAASG GERPNLKIDS TPVKPAGGGC
201 C

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_2i17, frame 3

SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B., N = 1, Score = 1023, P
= 2.7e-103

PIR:S06147 GTP-binding protein rab1B - rat, N = 1, Score = 1013, P =
3.2e-102

SWISSPROT:RAB1_DISOM RAS-RELATED PROTEIN ORAB-1., N = 1, Score = 967, P
= 2.4e-97

PIR:TVHUYP GTP-binding protein Rab1 - human, N = 1, Score = 966, P =
3e-97

>SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B.
Length = 201

HSPs:

Score = 1023 (153.5 bits), Expect = 2.7e-103, P = 2.7e-103
Identities = 197/201 (98%), Positives = 199/201 (99%)

Query: 1 MNPEYDYLFKLLLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ 60
MNPEYDYLFKLLLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ
Sbjct: 1 MNPEYDYLFKLLLLIGDSGVGKSCLLRFADDTYTESYISTIGVDKIRTIELDGKTIKIQ 60

Query: 61 IWDTAGQERFRTITSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIERYASENVNKLVLG 120
IWDTAGQERFRT+TSSYYRGAHGIIVVYDVTDQESYANVKQWLQEIERYASENVNKLVLG
Sbjct: 61 IWDTAGQERFRTVTSYYRGAHGIIVVYDVTDQESYANVKQWLQEIERYASENVNKLVLG 120

Query: 121 NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180
NKSDLTTKKVVDNTTAKEFADSLG+PFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
Sbjct: 121 NKSDLTTKKVVDNTTAKEFADSLGVPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG 180

Query: 181 GERPNLKIDSTPVKPAGGGCC 201
GERPNLKIDSTPVK A GGCC
Sbjct: 181 GERPNLKIDSTPVKSASGGCC 201

Pedant information for DKFZphfbr2_2i17, frame 3

Report for DKFZphfbr2_2i17.3

[LENGTH] 201

[MW] 22171.25
 [PI] 5.56
 [HOMOL] SWISSPROT:RB1B_RAT RAS-RELATED PROTEIN RAB-1B. 1e-112
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.08 organization of golgi [S. cerevisiae, YFL038c] 2e-77
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-57
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YER031c] 8e-46
 [FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae, YGL210w] 1e-44
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c] 1e-30
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 3e-25
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 9e-24
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 4e-23
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-17
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 1e-16
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 1e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 4e-10
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 9e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPL051w] 3e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL048c] 5e-05
 [BLOCKS] BL01019A ADP-ribosylation factors family proteins
 [BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
 [SCOP] dlpk_3.25.1.3.1 CH-p21 Ras protein [human (Homo sapiens) 2e-41
 [SCOP] dlguaa_3.25.1.3.10 Rap1A [Human (Homo sapiens) 5e-60
 [SCOP] dlrrga_3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattus 2e-30
 [SCOP] dlhura_3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Homo 2e-33
 [PIRKW] nucleus 1e-21
 [PIRKW] membrane trafficking 1e-110
 [PIRKW] oncogene 1e-25
 [PIRKW] endoplasmic reticulum 1e-105
 [PIRKW] phosphoprotein 1e-105
 [PIRKW] glycoprotein 3e-25
 [PIRKW] prenylated cysteine 1e-110
 [PIRKW] signal transduction 4e-23
 [PIRKW] transforming protein 1e-105
 [PIRKW] purine nucleotide binding 2e-24
 [PIRKW] alternative splicing 5e-26
 [PIRKW] P-loop 1e-110
 [PIRKW] lipoprotein 1e-110
 [PIRKW] proto-oncogene 3e-27
 [PIRKW] methylated carboxyl end 3e-27
 [PIRKW] hydrolase 7e-25
 [PIRKW] membrane protein 1e-105
 [PIRKW] GTP binding 1e-110
 [PIRKW] thiolester bond 5e-76
 [PIRKW] Golgi apparatus 1e-105
 [SUPFAM] ras transforming protein 1e-110
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 2
 [PROSITE] CK2_PHOSPHO_SITE 5
 [PROSITE] SIGMA54_INTERACT_1 1
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 4
 [PROSITE] ASN_GLYCOSYLATION 3
 [PFAM] Ras family (contains ATP/GTP binding P-loop)
 [KW] Alpha_Beta
 [KW] 3D

SEQ MNPEYDYLFKLLIGDSGVGKSCLLRFADDYTESYISTIGVDFKIRTIELDGKTIKIQ
221p-EEEEEEETTTCHHHHHHHHHHCCCCCCCCCTTTEEEE-EEEEETEEEEEE

SEQ IWDTAGQERFRTITSSYYRGAGHIIVVYDVTQESYANVKQWLQEI DRYASENVNKLVLG
221p- EEECTTTTTTCGGGHHHHHHHCCCEEEEEETTBHHHHHHHHHHHHHHHHHHHTTTCEEEEE

SEQ NKSDLTTKKVVDNTTAKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGAASG
221p- ETTTTCCTCC-CCCHHHHHHHHHHCCCCCEEEETTTTTTHHHHHHHHHHHHHHH.....

SEQ GERPNLKIDSTPVKPAAGGCC
221p-TTTTTTT

Prosite for DKFZphfbr2_2i17.3

PS000001	121->125	ASN_GLYCOSYLATION	PDOC000001
PS000001	133->137	ASN_GLYCOSYLATION	PDOC000001
PS000001	154->158	ASN_GLYCOSYLATION	PDOC000001
PS000002	17->21	GLYCOSAMINOGLYCAN	PDOC000002
PS000005	56->59	PKC_PHOSPHO_SITE	PDOC000005
PS000005	126->129	PKC_PHOSPHO_SITE	PDOC000005
PS000005	135->138	PKC_PHOSPHO_SITE	PDOC000005
PS000005	151->154	PKC_PHOSPHO_SITE	PDOC000005
PS000006	32->36	CK2_PHOSPHO_SITE	PDOC000006
PS000006	91->95	CK2_PHOSPHO_SITE	PDOC000006
PS000006	135->139	CK2_PHOSPHO_SITE	PDOC000006
PS000006	156->160	CK2_PHOSPHO_SITE	PDOC000006
PS000006	179->183	CK2_PHOSPHO_SITE	PDOC000006
PS000007	27->34	TYR_PHOSPHO_SITE	PDOC000007
PS000008	18->24	MYRISTYL	PDOC000008
PS000008	176->182	MYRISTYL	PDOC000008
PS000017	15->23	ATP_GTP_A	PDOC000017
PS00675	11->25	SIGMA54_INTERACT_1	PDOC00579

Pfam for DKFZphfbr2 2i17.3

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLIGDSGVGKSCLLIRFTQNEFnEeYIPTIGVDFYtKTIEIDGktIK KL+LIGDSGVGKSCLL+RF +++++E+YI+TIGVDF+++TIE+DGKTIK		
Query	10	KL LLIIGDSGVGKSCLLRFAFTTIGVDFTESYISTIGVDFKIRTIELDGKTIK	58
HMM	LQIWDTAGQERYRsMRPMYYRGAMGFMVLVDITNrqSFENiRnWweEiR LQIWDTAGQER+R++++YYRGA+G+++VYD+T+++S+ N+++W+++EI+R		
Query	59	LQIWDTAGQERFRITTSYYRGAHGIIIVYDVDTQESYANVQWLQEI DR	108
HMM	HCDrDENVPIMLVGNKCDLEDQRQVSTEEGQeFAREWGAIPFMETSAKTN +++ ENV ++LVGNK+DL +++V+ ++EFA+++G IPF+ETSAK++		
Query	109	YAS--ENVNKLVLGNKSDLTTKVVDNTTAKEFADSLG-IPFLETSAKNA	155
HMM	iNVEEAFMEiVReilqrMqe.q.NqteNinidQpsrnrk...rCCCIM* +NVE+AFM+++ EI+RKM+ +++E +N++ +S++ K ++C		
Query	156	TNVECAFMTMAAEI++RKGPGAASGGERPNLKI DSTPVKPAAGGCC--	201

DKFZphfbr2_2k19

group: brain derived

DKFZphfbr2_2k19 encodes a novel 303 amino acid protein with similarity to human KIAA0378 product.

The protein contains a leucine zipper, which can mediate protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to KIAA0378

encoded by the genomic clones HS147M19/HS608E8

Sequenced by Qiagen

Locus: unknown

Insert length: 1931 bp

Poly A stretch at pos. 1866, no polyadenylation signal found

```
1 GGGGGGGGCG CGCGGTGACA GCGCGGGGTT GCGGCGTGG GACCCAGGGG
51 GCGACAGAGG CAGCAGCAGC CCGAGGCCTG AGGAGAGGAG ACCGGCGGCG
101 GCGGCAATGC TGGAGACCTT TCGCGAGCGG CTGCTGAGCG TGCAGCAGGA
151 TTTCACCTCC GGGCTGAAGA CTTTAAGTGA CAAGTCAAGA GAAGCAAAAG
201 TGAAAAGCAA ACCCAGGACT GTTCCATTTT TGCCAAAGTA CTCTGCTGGA
251 TTAGAATTAC TTAGCAGGTA TGAGGATACA TGGGCTGCAC TTCACAGAAG
301 AGCCAAAGAC TGTGCAAGTG CTGGAGAGCT GGTGGATAGC GAGGTGGTCA
351 TGCTTTCTGC GCACTGGGAG AAGAAAAGA CAAGCCTCGT GGAGCTGCAA
401 GAGCAGCTCC AGCAGCTCCC AGCTTTAATC GCAGACTTAG AATCCATGAC
451 AGCAAAATCTG ACTCATTTAG AGGCGAGTTT TGAGGAGGTA GAGAACAACC
501 TGTGTGATCT GGAAGACTTA TGTGGGCAGT GTGAATTAGA AAGATGCAAA
551 CATATGCAGT CCCAGCAACT GGAGAATTAC AAGAAAATA AGAGGAAGGA
601 ACTTGAAGCC TTCAAAGCTG AACTAGATGC AGAGCACGCC CAGAAGGTCC
651 TGGAAATGGA GCACACCCAG CAAATGAAGC TGAAGGAGCG GCAGAAGTTT
701 TTTGAGGAAG CCTTCCAGCA GGACATGGAG CAGTACCTGT CCACTGGCTA
751 CCTGCAGATT GCAGAGCGGC GAGAGCCCAT AGGCAGCATG TCATCCATGG
801 AAGTGAACGT GGACATGCTG GAGCAGATGG TCCTGATGGA CATATCGGAC
851 CAGGAGGCCC TGGACGTCTT CCTGAACTCT GGAGGAGAAG AGAACACTGT
901 GCTGTCCCCC GCCTTAGGTA GGGTTGACAA ACTTGCAATTA GCTGAACCCG
951 GGCAGTATCG ATGCCACTCC CCTCCAAAGG TGAGACGTGA GAACCATCTG
1001 CCAGTCACTT ACGCATAAAC CCCCAGCTC ACAGCCAGCT CCTGGCTCCC
1051 TAACCCACAG GTTCCACACG GCTGTGTGGC AGCTGCAACA GTGGTGTGGT
1101 TCCGTGATGA ATTCTTCTCA AAGATTGAC ATGCTCCACT CCGGTAACCT
1151 TGGTGAGTTG AGAGCTTTCT TGTGTTGTTT CCCTCCTTTA CCATCCAGAA
1201 ATCCATTTGA GTCTGCTCCT TGTGGTTAAG GACTGGCGTT TGCAGGGAGG
1251 TGCGGACTCT CCTGCGGGGC TCACGGGAAA CTCTTCCCTC TTCGTGCGAC
1301 AGGCATTTAG GGGCGTGCCT GCCATGGGCA AAGCCATGGT GTGTGTTTCA
1351 CTCTTGGCCT GTGTTGTAAA CTTAGTTGCA CTTCAAGTCC TTTTCATCCCT
1401 TCACAAAATT TTGTTTCACA TTCATGCAGC AAATATGGGC TGAGGTGCCA
1451 GACCTGTACC TGGGCTTGGT GCGTTTCAAA TTTTCAGACCA GTTCTTTGGG
1501 CTGGGTCAAG GCAAAGCTCA GTCGTCCCAG CAGCACCTCA GCCATCTGTA
1551 GAAGGTCTTA CCATTACCAC GGTTCAGACT TCCTCTAAAC TTCTCACCCG
1601 CTTCTCTCTG CAATCTGTCA GAACGGTGTG ATCCTGGGGA AGAGAAGGAG
1651 CTTGGGTGCA TTTGCCCTCA TCCTGAGAAG GCCAGAATAC TGGAGACCCG
1701 CGTGAACCCCT CACCCAGAGT CAGGGGAAGA TTTAGAAACA GTGACACCTG
1751 CATATAGAAT TTTGATTCTT TGAAGAGCCT ATTTAGTTCC ATAAAATTGG
1801 AGAACTGCTG AAGGTGAGTA ATTCCGACTT TCTCAGCAGT GGTGTCTCTG
1851 AATTACTGCA AAGGGTAAAA AAAAAAATA AAAAAACTTA TCGATACCGT
1901 CGACCTCGAT GATGATGATG ATGATGTCGA C
```

BLAST Results

Entry HS147M19 from database EMBL:
Homo sapiens DNA sequence from PAC 147M19 on chromosome 6p22.1-22.3.
Contains an unknown gene, ESTs and GSSs.
Score = 5540, P = 4.1e-275, identities = 1114/1120
3 exons 592-1884

Entry HS608E8 from database EMBL:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 608E8
Score = 797, P = 1.2e-78, identities = 161/163

6 exons 1-592

Medline entries

90294724:

The involucrin gene of the gibbon: The middle region shared by the hominoids

Peptide information for frame 2

ORF from 107 bp to 1015 bp; peptide length: 303

Category: similarity to known protein

Classification: unset

Prosites motifs: LEUCINE_ZIPPER (97-119)

```

1 MLETLRERLL SVQQDFTSGL KTLSDKSREA KVKSKPRTVP FLPKYSAGLE
51 LLSRYEDTWA ALHRRRAKDC SAGELVDSEV VMLSAHWEKK KTSLEVELQEQ
101 LQQLPALIAD LESMTANLTH LEASFEEVEN NLLHLEDLCG QCELERCKHM
151 QSQQLENYKK NKRKELETFK AELDAEHAQK VLEMENTQOM KLKERQKFFE
201 EAFQQDMEQY LSTGYLQIAE RREPIGSMSS MEVNVDMLEQ MVLMDISDQE
251 ALDVFLNSGG EENTVLSPAL GRVDKLALAE PGQYRCHSPP KVRRENHLPV
301 TYA

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2k19, frame 2

TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds., N = 1, Score = 137, P = 4.8e-06

PIR:I37037 involucrin - common gibbon, N = 1, Score = 124, P = 7.4e-05

PIR:A57013 early endosome antigen 1 - human, N = 1, Score = 128, P = 9.5e-05

>TREMBL:HSAB2376_1 gene: "KIAA0378"; Human mRNA for KIAA0378 gene, partial cds.

Length = 808

HSPs:

Score = 137 (20.6 bits), Expect = 4.8e-06, P = 4.8e-06

Identities = 59/222 (26%), Positives = 103/222 (46%)

```

Query:      2 LETLRERLLSVQQDFTSGLKTL---SDKSREAKVKS-KPRTVPFLPKYSAGLELLSRYED 57
             L TL E L S ++   LK      D+ R +++S +   K +A   L+ E
Sbjct:    434 LATLEAL-SEKERITERLKEQREDDREERLEEIESFRKENKDLKEKVNALQAELETEKES 492

```

```

Query:      58 TWAALHRRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPALIADLESMTAN 117
             + L A ASAG DS++ L E+KK +L+ QL++ I D M
Sbjct:    493 SLIDLKEHASSLASAGLKRDSKLKSLEIAIEQKKECKSLEAQLKKAHN-IEDDSRMNPE 551

```

```

Query:      118 LTHLEASFEEVENNLLHLEDLCG--QCELERCKHMQSQQLENYKKNKRK---ELETFAE 172
             +++++ + D CG Q E++R + +++EN K +K K ELE+
Sbjct:    552 FAD---QIKQLDKEASYRDECGKAQAEVDRLEIL-KEVENEKNDKDKKIAELESLETLR 607

```

```

Query:      173 LDAEHAQKVLEMEHTQOMKLKERQKFFEEAFQQDMEQYLSLSTGYLQIAE 220
             + +KV ++H QQ++ K+ + EE +++ ++ +LQI E
Sbjct:    608 HMKDQNKVAVNLKHNQQLKKNNAQLLEEVRRREDSMADNSQHLQIEE 655

```

Score = 100 (15.0 bits), Expect = 6.2e-02, P = 6.0e-02

Identities = 44/156 (28%), Positives = 76/156 (48%)

```

Query:      57 DTWAALHRRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQQLPAL- IADLESMT 115
             D A+ +R +C A VD + +L E +K + +L+ L + D
Sbjct:    560 DKEASYR--DECGKAQAEVDRLEILK-EVENEKNDKDKKIAELESLETLRHMKDQNKV 616

```

```

Query:      116 ANLTHLEASFEEVENNLLHLEDLCGQCE--LERCKHMQSQQLENYKKNKRKELETFKAEL 173

```

ANL H + E+ +N L LE++ + + + +H+Q ++L N + R+EL+ KA L
 Sbjct: 617 ANLKHNQ-QLEKKKNAQL-LEEVRREDSDMADNSQHLQIEELMNALEKTRQELDATKARL 674
 Query: 174 DAEHAQKVLEME-HTQQMKLKERQKFFEEAFQQDMEQYLS 212
 A Q + E E H +++ ER+K EE + E L+
 Sbjct: 675 -ASTQQSLAEKEAHLANLRI-ERRKQLEEILEMKQEALLA 712

Pedant information for DKFZphfbr2_2k19, frame 2

Report for DKFZphfbr2_2k19.2

```
[LENGTH]      303
[MW]           34814.78
[pI]           5.23
[PROSITE]      LEUCINE_ZIPPER 1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      3.63 %
[KW]           COILED_COIL        14.52 %
```

SEQ MLETLRERLLSVQQDFTSGLKTLSDKSREAKVKSPPRTVPFLPKYSAGLELLSRYEDTWA
 SEG
 PRD ccchhhhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhhhhhchhh
 COILS

SEQ ALHRRAKDCASAGELVDSEVVMLSAHWEKKKTSLEVELQEQLPALIADLESMTANLTH
 SEGxxxxxxxxx.....
 PRD hhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 COILSCCCCCCCCCCCCCCCCCCCC

SEQ LEASFEEVENNLHLEDLCGQCELERCKHMQSQOLENYKKNRKELETFAELDAEHAQK
 SEG
 PRD hhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
 COILS CCCCCCCCCCCCCCCCCC.....

SEQ VLEMEHTQQMKLKERQKFFEEAFQQDMEQYLSYGYLQIAERREPIGSMSSMEVNVDMLEQ
 SEG
 PRD hhh
 COILS

SEQ MVLMDISDQEALDVFLNSGGEENTVLSPALGRVDKLALAEPGQYRCHSPPKVRRENHLPV
 SEG
 PRD hhh
 COILS

SEQ TYA
 SEG ...
 PRD CCC
 COILS ...

Prosite for DKFZphfbr2_2k19.2

PS00029 97->119 LEUCINE_ZIPPER PDOC00029

(No Pfam data available for DKFZphfbr2_2k19.2)

DKFZphfbr2_2k14

group: cell cycle

DKFZphfbr2_2k14 encodes a novel 335 amino acid protein with strong similarity to rattus rattus IAG2 "implantation-associated protein" and the human N33 tumour-suppressor gene.

Tumour-suppressor genes are known to be involved in the control of cell growth and division, interacting with proteins which control the cell cycle. The N33 gene is significantly methylated in tumour cells, a mechanism by which tumor-suppressor genes are inactivated in cancer. In addition, the novel protein contains a RGD cell attachment site. Therefore the novel protein is a new putative tumour-suppressor gene.

The new protein can find application in modulating/blocking the cell cycle and in the therapy of tumours.

strong similarity to human N33 tumor suppressor gene

complete cDNA, complete cds, EST hits,
potential start at Bp 30 matches kozak consensus ANCatgG
potential transmembran protein (4 TM)
similarity to yeast OST3p (oligosaccharyltransferase gamma chain)

Sequenced by Qiagen

Locus: unknown

Insert length: 2241 bp
Poly A stretch at pos. 2221, no polyadenylation signal found

```

1  TGGGACTTAT AGAAGGGAGA GGAGCGAACA TGGCAGCGCG TTGGCGGTTT
51  TGGTGTGTCT CTGTGACCAT GGTGGTGGCG CTGCTCATCG TTTGCGACGT
101  TCCCTCAGCC TCTGCCCAAA GAAAGAAGGA GATGGTGTTA TCAGAAAAGG
151  TTAGTCAGCT GATGGAATGG ACTAACAAA GACCTGTAAT AAGAATGAAT
201  GGAGACAAGT TCCGTGCGCT TGTGAAAGCC CCACCGAGAA ATTACTCCGT
251  TATCGTCATG TTCACTGCTC TCCAACGCA TAGACAGTGT GTCGTTTGCA
301  AGCAAGCTGA TGAAGAATTC CAGATCCTGG CAAACTCCTG GCGATACTCC
351  AGTGCATTCA CCAACAGGAT ATTTTGTGCC ATGGTGGATT TTGATGAAGG
401  CTCTGATGTA TTTCAGATGC TAAACATGAA TTCAGCTCCA ACTTTCATCA
451  ACTTTCCTGC AAAAGGGAAA CCCAAACGGG GTGATACATA TGAGTTACAG
501  GTGCGGGGTT TTTCAGCTGA GCAGATTGCC CGGTGGATCG CCGACAGAAC
551  TGATGTCAAT ATTAGAGTGA TTAGACCCCC AAATTATGCT GGTCCCCTTA
601  TGTGGGATT GCTTTTGGCT GTTATTGGTG GACTTGTGTA TCTTCGAAGA
651  CTTAATATGG AATTTCTCTT TAATAAACT GGATGGGCTT TTGAGCTTTT
701  GTGTTTGTG CTTGCTATGA CATCTGGTCA AATGTGGAAC CATATAAGAG
751  GACCACCATA TGCCCATAG AATCCCAACA CGGGACATGT GAATTATATC
801  CATGGAAGCA GTCAAGCCCA GTTTGTAGCT GAAACACACA TTGTTCTTCT
851  GTTTAATGGT GGAGTTACCT TAGGAATGGT GCTTTTGTGT GAAGCTGCTA
901  CCTCTGACAT GGATATTGGA AAGCGAAAGA TAATGTGTGT GGCTGGTATT
951  GGACTTGTG TATTATTCTT CAGTTGGATG CTCCTCTATT TTAGATCTAA
1001  ATATCATGGC TACCCATACA GCTTTCTGAT GAGTTAAAAA GGTCCCAGAG
1051  ATATATAGAC ACTGGAGTAC TGGAAATTGA AAAACGAAAA TCGTGTGTGT
1101  TTGAAAAGAA GAATGCAACT TGTATATCT GTATTACCTC TTTTTCCTAA
1151  GTGATTTAAA TAGTTAATCA TTTAACCAA GAAGATGTGT AGTGCCTTAA
1201  CAAGCAATCC TCTGTCAAAA TCTGAGGTAT TTGAAAATAA TTATCCTCTT
1251  AACCTTCTCT TCCCAGTGAA CTTTATGGAA CATTTAATTT AGTACAATTA
1301  AGTATATTAT AAAAAATTGA AAACACTACT TTTGTTTGTG TTAGAACAAA
1351  CCTCAAAACT ACTTTAGTTA ACTTGGTCTA CTGATCTTAT ATTGCCTTAT
1401  GCAAGATGGG GGAAGTAAG TCCTGACCAG GTGTTCCAC ATATGCCTGT
1451  TACAGATAAC TACATTAGGA ATTCATTCTT AGCTTCTTCA TCTTTGTGTG
1501  GATGTGTATA CTTTACGCAT CTTTCCTTTT GAGTAGAGAA ATTATGTGTG
1551  TCATGTGGTC TTCTGAAAAT GGAACACCAT TCTTCAGAGC ACACGTCTAG
1601  CCCACAGCAA GACAGTTGTT TCTCCTCCTC CTTGCATATT TCCTACTGCG
1651  CTCAGCCTG AGTGATAGAG TGAGACTCTG TCTCAAAAAA AAAGTATCTC
1701  TAAATACAGG ATTATAATTT CTGCTTGAGT ATGGTGTTAA CTACCTTGTA
1751  TTTAGAAAGA TTTAGATTTC ATTCATCTCT CTTAGTTTTC TTTAAGGTG
1801  ACCCATCTGT GATAAAAAATA TAGCTTAGTG CTAAATCAG TGTAACCTAT
1851  ACATGGCCTA AAATGTTTCT ACAAATTAGA GTTTGTCAC TATTCCATTT
1901  GTACCTAAGA GAAAAATAGG CTCAGTTAGA AAAGGACTCC CTGGCCAGGC
1951  CGAGTGACTT ACGCCTGTAA TCTCAGCACT TTGGGAGGCC AAGGCAGGCA
2001  GATCAGCAGG TCAGGAGTTC GAGACCATCC TGGCCAACAT GGTGAAACCC
2051  CGTCTCTACT AAAAAATATA AAATTAGCTG GGTGTGGTGG CAGGAGCCTG
2101  TAATCCCAGC TGCACAGGAG GCTGAGGCAC GAGAATCACT TGAACCTCAGG
2151  AGATGGAGGT TTCAGTGAGC CGAGATCAGC CCACTGCAC TACAGCCTGGC
2201  AACAGAGCGA GACTCCATCT CAAAAAATAA AAAAAAATAA A

```

BLAST Results

No BLAST result

Medline entries

96299740:
Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22.

97243398:
Tumour-suppressor genes in prostatic oncogenesis: a positional approach.

98334474:
Concordant methylation of the ER and N33 genes in glioblastoma multiforme.

Peptide information for frame 3

ORF from 30 bp to 1034 bp; peptide length: 335
Category: strong similarity to known protein

```

1 MAARWRFCV SVTMVVALLI VCDVPSASQ RKKEMVLSEK VSQLEWNTNK
51 RPVIRMNGDK FRRLVKAPPR NYSVIVMFTA LQLHRQCVCV KQADEEFQIL
101 ANSWRYSSAF TNRIFFAMVD FDEGSDVFQM LNMNSAPTFI NFPAGKPKR
151 GDTYELQVRG FSAEQIARWI ADRTDVNIRV IRPPNYAGPL MLGLLAVIG
201 GLVYLRRSNM EFLFNKTGWA FAALCFVLAM TSGQMWNHIR GPPYAHKNPH
251 TGHVNYIHGS SQAQFVAETH IVLLENGGVT LGMVLLCEAA TSDMDIGKRK
301 IMCVAGIGLV VLEFFSWMLSI FRSKYHGYPY SFLMS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_2k14, frame 3

TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds., N = 1, Score = 1560, P = 3.4e-160

PIR:G02297 gene N33 protein - human, N = 1, Score = 1256, P = 5.6e-128

TREMBL:HSN33S11_1 gene: "N33"; product: "N33 protein form 2"; Human N33 protein form 2 (N33) gene, exon 11 and complete cds., N = 1, Score = 1252, P = 1.5e-127

>TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein"; Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds. Length = 308

HSPs:

Score = 1560 (234.1 bits), Expect = 3.4e-160, P = 3.4e-160
Identities = 295/307 (96%), Positives = 299/307 (97%)

```

Query:   29 AQRKKEMVLSEKVSQLEWNTNRPVIRMNGDKFRRLVKAPPRNYSVIVMFTALQLHRQCV 88
          AQRKKE VL EKV QLEWNTN+RPVIRMNGDKFR LVKAPPRNYSVIVMFTALQLHRQCV
Sbjct:   2 AQRKKEVLVEKVIQLEWNTNQRPVIRMNGDKFRPLVKAPPRNYSVIVMFTALQLHRQCV 61

Query:   89 VCKQADEEFQILANSWRYSSAF TNRIFFAMVD FDEGSDVFQMLNMNSAPTFINFPAGKGP 148
          VCKQADEEFQILAN WRYSSAF TNRIFFAMVD FDEGSDVFQMLNMNSAPTFINFP KGKP
Sbjct:   62 VCKQADEEFQILANFWRYSSAF TNRIFFAMVD FDEGSDVFQMLNMNSAPTFINFPKGKP 121

Query:   149 KRGDYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 208
          KR DTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS
Sbjct:   122 KRADTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNYAGPLMLGLLAVIGGLVYLRRS 181

Query:   209 NMEFLFNKTGWAFALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE 268
          NMEFLFNKTGWAFALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE

```

Sbjct:	182	NMEFLFNKGTGWAFALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAE	241
Query:	269	THIVLLFNGGVTGLGMVLLCEAATSDMDIGRKIMCVAGIGLVVLFSSWMLSIFRSKYHGY	328
		THIVLLFNGGVTGLGMVLLCEAA SDMDIGRR+MC+AGIGLVVLFSSWMLSIFRSKYHGY	
Sbjct:	242	THIVLLFNGGVTGLGMVLLCEAAASDMDIGRRMMCIAGIGLVVLFSSWMLSIFRSKYHGY	301
Query:	329	PYSFLMS 335	
		PYSFLMS	
Sbjct:	302	PYSFLMS 308	

Pedant information for DKFZphfbr2 2k14, frame 3

Report for DKFZphfbr2 2k14.3

```

[LENGTH]          335
[MW]               38036.83
[pI]               9.68
[HOMOL]            TREMBL:RNAF8554_1 gene: "IAG2"; product: "implantation-associated protein";
Rattus norvegicus implantation-associated protein (IAG2) mRNA, partial cds. 1e-161
[FUNCAT]           30.07 organization of endoplasmatic reticulum [S. cerevisiae, YOR085w]
4e-14
[FUNCAT]           06.07 protein modification (glycolsylation, acylation, myristylation,
palmitylation, farnesylation and processing) [S. cerevisiae, YOR085w] 4e-14
[FUNCAT]           01.05.01 carbohydrate utilization [S. cerevisiae, YOR085w] 4e-14
[EC]               2.4.1.119 Dolichyl--diphosphooligosaccharide--protein glycosyltransferase 1e-12

[PIRKW]            glycosyltransferase 1e-12
[PIRKW]            transmembrane protein 6e-69
[PIRKW]            hexosyltransferase 1e-12
[PROSITE]          RGD      1
[PROSITE]          MYRISTYL      4
[PROSITE]          AMIDATION      1
[PROSITE]          CK2_PHOSPHO_SITE      2
[PROSITE]          PKC_PHOSPHO_SITE      4
[PROSITE]          ASN_GLYCOSYLATION      2
[KW]               SIGNAL PEPTIDE 30
[KW]               TRANSMEMBRANE 4
[KW]               LOW COMPLEXITY      5.97 %

```

```

SEQ      MAARWRFWCVSVTMVVALLIVCDVPSASAQRKKEMVLSEKVSQLMEWTKRNPVIRMNGDK
SEG
PRD      cccceeeeeehhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhhccceeeecccc
MEM      .....

SEQ      FRRLVKAPPNYSVIVMFTALQLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFFAMVD
SEG
PRD      ceeeeecccccceeeehhhhhhccceeeehhhhhhhhhhhhhccccccccceeeec
MEM      .....

SEQ      FDEGSDVFQMLNMNSAPTFINFPAGKPKRGDTYELQVRGFSAEQIARIWIADRTDVNIRV
SEG
PRD      cccccceeeecccccceeeecccccccccceeeecchhhhhhhhhhhheeeee
MEM      .....M

SEQ      IRPPNYAGPLMLGLLLAVIGLVLYLRRSNEMFLFNKTGWAFALCFVLAMTSGQMWNHIR
SEG      ....xxxxxxxxxxxxxxxxxxxxxx.
PRD      eccccchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceec
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM. ....MMMMMMMMMMMMMMMMMMMMMMMMMM...

SEQ      GPPYAHKNPHTGHVNYIHGSSQAQFVAETHIVLLFNGVTLGMLVLLCEAATSDMDIGKRK
SEG
PRD      cccccccccccccceeeccchhhhhhhheeeecchhhhhhhhhhhhhcccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      IMCVAGIGLVVLFSSWMLSIFRSKYHGYPSFLMS
SEG
PRD      eeeccccceeeehhhhhhhhhhhccccc
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

```

Prosite for DKFZphfbr2 2k14.3

PS00001	71->75	ASN_GLYCOSYLATION	PDOC00001
PS00001	215->219	ASN_GLYCOSYLATION	PDOC00001
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005

WO 01/12659

PCT/IB00/01496

PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00008	193->199	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	278->284	MYRISTYL	PDOC00008
PS00009	296->300	AMIDATION	PDOC00009
PS00016	150->153	RGD	PDOC00016

(No Pfam data available for DKFZphfbr2_2k14.3)

DKFZphfbr2_3c18

group: nucleic acid management

DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicase and RNA-dependent ATPase
from the DEAD box family
group helicases

Summary DKFZphfbr2_3c18 encodes a novel 448 amino acid protein with
similarity to DEAD-box subfamily ATP-dependent RNA helicases.
Deletion of the yeast homologue DBP5 is lethal.

strong similarity to RNA helicase and RNA-dependent ATPase from the
DEAD box family

complete cDNA, EST hits
complete cds ATG at Bp 109

Sequenced by AGOWA

Locus: /map="87.50 cR from top of Chr16 linkage group"

Insert length: 1713 bp
Poly A stretch at pos. 1696, no polyadenylation signal found

```

1 TGGGGTAGTG GGGCTGGAGC AGAGCCTGCC GCGAACCCCG GGAGCCACAG
51 ATCCCTCGTG CCATCCCTCG AATCCACCAG CACGAGCGTC CCACCCGCGC
101 CTGGGACCAT GGCCACTGAC TCATGGGCCC TGGCGGTGGA CGAGCAGGAA
151 GCTGCGGCTG AGTCGTTGAG CAACTTGCAT CTTAAGGAAG AGAAAATCAA
201 ACCAGATACC AATGGTGCTG TTGTCAAGAC CAATGCCAAT GCAGAGAAGA
251 CAGATGAAGA AGAGAAAGAG GACAGAGCTG CCCAGTCCTT ACTCAACAAG
301 CTGATCAGAA GCAACCTTGT TGATAACACA AACCAAGTGG AAGTCCTGCA
351 GCGGGATCCA AACTCCCTCT TGTACTCGGT GAAGTCTTTT GAAGAGCTTC
401 GGCTCCCAAC GAACTTAATT GCCCAATCTC AGTCTGGTAC TGGTAAAAA
451 GCTGCCTTCG TGCTGGCCAT GCTTAGCCAA GTAGAAGCTG CAAACAAAAT
501 CCCCCAGTGT CTATGTCTCT CCCCACGTA TGAGCTCGCC CTCCAAACAG
551 GAAAAGTGAT TGAACAAATG GGCAAAATTT ACCCTGAAGT GAAGCTAGCT
601 TATGCTGTTC GAGGCAATAA ATTGGAAAGA GGCCAGAAGA TCAGTGAGCA
651 GATTGTCATT GGCACCCCTG GGAAGTGTCT GGACTGGTGC TCCAAGCTCA
701 AGTTTCATTG TCCCAAGAAA ATCAAGGTGT TTGTTCTGGA TGAGGCTGAT
751 GTCATGATAG CCACTCAGGG CCACCAAGAT CAGAGCATCC GCATCCAGAG
801 GATGCTGCCC AGGAAGTACC AGATGCTGCT TTTCTCCGCC ACCTTTGAAG
851 ACTCTGTGTG GAAGTTTGCC CAGAAAGTGG TCCAGAGCCC AAACGTTATC
901 AAAGTGAAGC GTGAGGAAGA GACCTGGAC ACCATCAAGC AGTACTATGT
951 CCTGTGCAGC AGCAGAGACG AGAAGTTCCA GGCCTTGTGT AACCTCTACG
1001 GGGCCATCAC CATTGCTCAA GCCATGATCT TCTGCCATAC TCGCAAAACA
1051 GCTAGTTGGC TGGCAGCAGA GCTCTCAAAA GAAGGCCACC AGGTGGCTCT
1101 GCTGAGTGGG GAGATGATGG TGGACAGAG GGCTGCAGTG ATTGAGCGCT
1151 TCCGAGAGGG CAAAGAGAAG GTTTTGGTGA CCACCAACGT GTGTGCCCGC
1201 GGCATTGATG TTGAACAAGT GTCTGTCGTC ATCAACTTTG ATCTTCCCGT
1251 GGACAAGGAC GGAATCCTG ACAATGAGAC CTACCTGCAC CGGATCGGGC
1301 GCACGGGCGG CTTTGGCAAG AGGGGCTGG CAGTGAACAT GGTGGACAGC
1351 AAGCACAGCA TGAACATCCT GAACAGAATC CAGGAGCATT TTAATAAGAA
1401 GATAGAAAGA TTGGACACAG ATGATTGGGA CGAGATTGAG AAAATAGCCA
1451 ACTGAGAAGC TCCACCAGCC ACTGATGCCA GCCCTGGCAC TGCCCCTGCA
1501 CAGGAGACAA GTGCGTTCAG GGCACAGGCC CCGACATCAC CCCAAGGACA
1551 ACGGCACAAG TAGAGAGAAA CTACCTACCT CACTTCAAAAT TATGTTTGG
1601 CTTGACAAAA ATGTATGCAA ATGATGGGGG ATGGTAGAAA AAAATTATT
1651 ACACAACCTT GGAAGATTAG GCATGAATAC ACAGAGATTT ACCTTTAAAA
1701 AAAAAAAAAA AAA

```

BLAST Results

Entry G36496 from database EMBL:
 SHGC-53094 Human Homo sapiens STS cDNA.
 Length = 459
 Minus Strand HSPs:
 Score = 1693 (254.0 bits), Expect = 2.8e-70, P = 2.8e-70
 Identities = 369/387 (95%), Positives = 369/387 (95%)

Entry G44014 from database EMBLNEW:
 WIAF-3643-STS Human THudson SANGER Homo sapiens STS genomic, sequence
 tagged site.
 Score = 901, P = 2.3e-35, identities = 183/185

Medline entries

94192995:
 Gene 1994 Mar 25;140(2):171-177
 Mouse erythroid cells express multiple putative RNA helicase genes
 exhibiting
 high sequence conservation from yeast to mammals.

Peptide information for frame 1

ORF from 109 bp to 1452 bp; peptide length: 448
 Category: strong similarity to known protein

```

1 MATDSWALAV DEQEAAASL SNLHLKEEKI KPDTNGAVVK TNANAECTDE
51 EEKEDRAAQS LLNKLIRSNL VDNTNQVEVL QRDPNSPLYS VKSFEELRLP
101 QNLIAQSQSG TGKTAAFVLA MLQVEPANK YPQCLCLSPT YELALQTGKV
151 IEQMGKFYPE LKLAYAVRGN KLERGQKISE QIVIGTPGTV LDWCSKLKFI
201 DPKKIKVFVL DEADVMIATQ GHQDQSIRIQ RMLPRNCQML LFSATFEDSV
251 WKFAQKVVPD PNVIKLRKEE ETLDTIKQYY VLCSSRDEKF QALCNLYGAI
301 TIAQAMIFCH TRKTASWLA ELKKEGHQVA LLSGEMMVEQ RAAVIERFRE
351 GKEKVLVTTN VCARGIDVEQ VSVVINFDLP VDKDGNPDNE TYLHRIGRTG
401 RFGKRGGLAVN MVDKSHSMNI LNRIQEHFNK KIERLDTDDL DEIEKIAN

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3c18, frame 1

PIR:I49731 RNA helicase - mouse, N = 2, Score = 1758, P = 3.8e-223

TREMBL:AF005239_1 gene: "Dbp80"; product: "DEAD-box helicase";
 Drosophila melanogaster DEAD-box helicase (Dbp80) mRNA, complete cds.,
 N = 2, Score = 1142, P = 1.8e-125

SWISSPROT:YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06., N =
 2, Score = 911, P = 5.5e-103

PIR:S66920 probable RNA helicase CA5/6 - yeast (Saccharomyces
 cerevisiae), N = 2, Score = 887, P = 1.9e-98

>PIR:I49731 RNA helicase - mouse
 Length = 478

HSPs:

Score = 1758 (263.8 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 338/349 (96%), Positives = 349/349 (100%)

```

Query: 100 PQNLIAQSQSGTGKTAAFVLA MLQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYP 159
      PQNLIAQSQSGTGKTAAFVLA MLQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYP
Sbjct: 130 PQNLIAQSQSGTGKTAAFVLA MLQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFHP 189

Query: 160 ELKLAYAVRGNKLERGQKISEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT 219
      ELKLAYAVRGNKLERGQKISEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT
Sbjct: 190 ELKLAYAVRGNKLERGQKISEQIVIGTPGTVDWCSKLKFDPPKKIKVFVLDEADVMIAT 249

Query: 220 QGHQDQSIRIQRLPRNCQMLLFSATFEDSVWKFAQKVVPDPNVIKLRKEETLDTIKQY 279

```

QGHQDQSIRIQR++PRNCQMLLFSAFEDSVWKFQKVVDPDN+IKLKREEETLOTIKQY
 Sbjct: 250 QGHQDQSIRIQRIVPRNCQMLLFSAFEDSVWKFQKVVDPDNIIKLKREEETLOTIKQY 309
 Query: 280 YVLCSSRDEKFQALCNLYGAIITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 339
 YVLC++R+EKFOALCNLYGAIITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE
 Sbjct: 310 YVLCNNREEKFQALCNLYGAIITIAQAMIFCHTRKTASWLAELSKEGHQVALLSGEMMVE 369
 Query: 340 QRAAVIERFREGKEKVLVTNVCARGIDVEQSVVINFDLPVDKDGNDPNETYLHRIGRT 399
 QRAAVIERFREGKEKVLVTNVCARGIDVEQSVVINFDLPVDKDGNDPNETYLHRIGRT
 Sbjct: 370 QRAAVIERFREGKEKVLVTNVCARGIDVEQSVVINFDLPVDKDGNDPNETYLHRIGRT 429
 Query: 400 GRFGKRGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 448
 GRFGKRGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN
 Sbjct: 430 GRFGKRGLAVNMVDSKHSNMILNRIQEHFNKKIERLDTDDLDEIEKIAN 478
 Score = 419 (62.9 bits), Expect = 3.8e-223, Sum P(2) = 3.8e-223
 Identities = 94/136 (69%), Positives = 104/136 (76%)
 Query: 1 MATDSWALAVDEQEAAESLSNLHLKEEKIKPDTNGAVVKTNANAECTDEEEKEDRAAQS 60
 MATDSWALAVDEQEAA +S+S+L +KEEK K DTNG V+KT+ AEKT+EEEEKEDRAAQS
 Sbjct: 1 MATDSWALAVDEQEAAVKSMSLSLQIKKEAKSDTNG-VIKTSTTAEKTEEEKEDRAAQS 59
 Query: 61 LLNKLIRSNLVDNTNQVEVLQRPNSPLYSVKSFEELRL-PQNL---IAQSQSGTGKTA 116
 LLNKLIRSNLVDNTNQVEVLQRPNSPLYSVKSFEELRL PQ L A + K
 Sbjct: 60 LLNKLIRSNLVDNTNQVEVLQRPSSPLYSVKSFEELRLKQLLQGVYAMGFNRPSKIQE 119
 Query: 117 FVLAMLSQVEPANKYPQ 133
 L M+ P N Q
 Sbjct: 120 NALPMLAEPPQNLIQ 136

Pedant information for DKFZphfbr2_3c18, frame 1

Report for DKFZphfbr2_3c18.1

[LENGTH] 448
 [MW] 50490.07
 [PI] 5.83
 [HOMOL] PIR:I49731 RNA helicase - mouse 0.0
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 1e-102
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR021w] 2e-65
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YJL138c] 1e-63
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 2e-49
 [FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 9e-48
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YDL084w] 1e-43
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 3e-39
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-35
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 9e-27
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 8e-26
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 1e-23
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 9e-08
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 1e-05
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIR002c] 7e-04
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-64
 [PIRKW] RNA binding 1e-64
 [PIRKW] DEAD box 4e-64
 [PIRKW] transmembrane protein 3e-22
 [PIRKW] DNA binding 2e-32
 [PIRKW] ATP 1e-101
 [PIRKW] purine nucleotide binding 4e-64
 [PIRKW] P-loop 1e-101
 [PIRKW] hydrolase 4e-43
 [PIRKW] protein biosynthesis 1e-64
 [PIRKW] ATP binding 2e-35
 [SUPFAM] WW repeat homology 3e-29
 [SUPFAM] translation initiation factor eIF-4A 1e-64
 [SUPFAM] DEAD/H box helicase homology 1e-101
 [SUPFAM] DNA helicase recG 2e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-101
 [SUPFAM] ATP-dependent RNA helicase DBP1 9e-33

SEQ	MATDSWALAVDEQEAASLSNLHLKEEIKPDTNGAVKTNANAETDEEEKEDRAAQS
PRD	ccchhhhhhhhhhhhhhhhhccchhhhhhhhhccccceeeehhhhhhhhhhhhhhhhh
SEQ	LLNKLIRSNLVDNTNQVEVLQRDPNSPLYSVKSFEELRLPQNLIASQSGTGKTAAFVLA
PRD	hhhhhhhhhhccccceeeeeeccccceeehhhhhhhhhhccceeeccccccchhhhh
SEQ	MLSQVEPANKYQPCLECSPTYELALQTGKVI EQMGKFYPELKLAVAYRGNKLERGQKISE
PRD	hhhhhhhhhhccccceeeccccchhhhhhhhhhhccccceeeccccchhhhhhhhh
SEQ	QIVIGTPTGTVLDWC SKLKFIDPKKIKVFVLDEADVMIATQGHQDQSIRIQRMLPRNCQML
PRD	eeeeeccccchhhhhhhhhhhccccceeeecchhhhhhhccchhhhhhhhhhhcccccee
SEQ	LFSATFEDSVWKFAQKVPDPNVIKLKREEETLDTIKQYVVLCSRRDEKFOALCNLYGAI
PRD	eeccccchhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhch
SEQ	TIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTTN
PRD	hhhhhhheecchhhhhhhhhhhhhccccceeeccccchhhhhhhhhhhccccceeeeee
SEQ	VCARGIDVEQSVVINFOLPVDKDGPNPNETYLHRIGRTRGFRGKRLAVNMVDSKSHSMNI
PRD	ccccccccceeeeeeccccccccccccccccceeeccccccccccccceeeccccchhh
SEQ	LNRIQEHFNKKIERLDTDDLDEIEKIAN
PRD	hhhhhhhhhhhhccccccccccccchhhhhcc

PS000001	389->393	ASN_GLYCOSYLATION	PDOC000001
PS000002	109->113	GLYCOSAMINOGLYCAN	PDOC000002
PS000005	90->93	PKC_PHOSPHO_SITE	PDOC000005
PS000005	111->114	PKC_PHOSPHO_SITE	PDOC000005
PS000005	147->150	PKC_PHOSPHO_SITE	PDOC000005
PS000005	226->229	PKC_PHOSPHO_SITE	PDOC000005
PS000005	275->278	PKC_PHOSPHO_SITE	PDOC000005
PS000005	284->287	PKC_PHOSPHO_SITE	PDOC000005
PS000005	311->314	PKC_PHOSPHO_SITE	PDOC000005
PS000005	399->402	PKC_PHOSPHO_SITE	PDOC000005
PS000006	48->52	CK2_PHOSPHO_SITE	PDOC000006
PS000006	93->97	CK2_PHOSPHO_SITE	PDOC000006
PS000006	123->127	CK2_PHOSPHO_SITE	PDOC000006
PS000006	189->193	CK2_PHOSPHO_SITE	PDOC000006
PS000006	245->249	CK2_PHOSPHO_SITE	PDOC000006
PS000006	284->288	CK2_PHOSPHO_SITE	PDOC000006
PS000008	110->116	MYRISTYL	PDOC000008
PS000008	175->181	MYRISTYL	PDOC000008
PS000008	185->191	MYRISTYL	PDOC000008
PS000008	385->391	MYRISTYL	PDOC000008
PS000008	406->412	MYRISTYL	PDOC000008
PS000009	402->406	AMIDATION	PDOC000009

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPpWILRnIyeMGFEkPTPIQQqAIPiILeG...RDVMAcAQTSgSK		
	++ ++ +N ++	P	E+ ++A++G+G+GK
Query	65	LIRSLNVdNTNNQVLEVRQDPSNPLYSVKSFEELRLPQNLIASQSGTGK	113
HMM	TAAFLIPMLQHIWdPwppqPQdPrALILAPTRELAMQIEEcRkFgkHM		
	TAAFL+ ML+++ +	+ PQ	+L L+PT ELA+Q+ +++++GK++
Query	114	TAAFLVLAmsLsqVEpAN--KYQP---CLCLSPTYELALQTKGKVIeQMGKFY	158
HMM	ngIRimCIYGGtnMRdQMRmLeRgPPHIViATPGRLIDHIER.gtldLdr		
	++ ++ ++ ++	+++ ++ ++ ++	+IVI+TPG ++D + +D ++

Query 159 PELKLAYAVR---GNKLERGQKISEQIVIGTPGTVLDWCSKLFIDPKK 204
HMM IeMLVMDEADRMLD.MGFIDQIRrIMrqIPmpwNRQTMMFSATMPdeIqE
I+++V+DEAD M+ +G +DQ RI R++P +N Q ++FSAT+ D++ +
Query 205 IKVFLVDEADVMIATQGHQDQSIRIQRMPL--RNCQMLLFSATFEDSVWK 252
HMM LARrFMRNPiRInIdMdELTtnEnIkQwYiyVerEMWkfdcLcrLle*
+A ++ +P I ++++E T++ +IKQ+Y+ + + ++KF +LC+L++
Query 253 FAQKVVPDPNVIKLRKEETLD-TIKQYYVLCSSRDEKFAQLCNLYG 298

HMM_NAME Helicases conserved C-terminal domain

HMM *EileeWLknlGirvmYIHGdMpQeERdeIMddFnnGEynVLicTDVggr
+L+ +L+++G +V+ + G M+ E+R ++++F++G+ +VL++T+V +R
Query 316 SWLAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTNNVCAR 364
HMM GIDIPdVNHVINYDM...PWNPEq..YIQRIGRTgRIG*
GID+++V++VIN+D+ + NP++ Y++RIGRTGR+G
Query 365 GIDVEQVSVINFDLPVDKDGNDNETYLHRIGRTGRFG 403

Medline

PMID: 10322435

"Unwinding RNA in : DEAD-box proteins and related families." de la Cruz J, Kressler D, Linder
P

DKFZphfbr2_3f16

group: brain derived

DKFZphfbr2_3f16 encodes a novel 127 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1514 bp

Poly A stretch at pos. 1454, polyadenylation signal at pos. 1434

```
1 GGGGGGACTG GAGAAGGGAG GCGGCGGGCG AAGCGCACGT CGAGCGGGGG
51 AGCGGCGCTG CCTGTGGAGA TCCGCGGAGG CCGACAGGAT TCGTTGGCTG
101 CCGTCCCCGC TGCTGTGCAT TGGGTAAAAA ACGACAACCA ACATCAGCCA
151 TGAAAGATCC AAGTCGCAGC AGTACTAGCC CAAGCATCAT CAATGAAGAT
201 GTGATTATTA ACGGTCATTC TCATGAAGAT GACAATCCAT TTGCAGAGTA
251 CATGTGGATG GAAAAATGAG AAGAATTCAA CAGACAAATA GAAGAGGAGT
301 TATGGGAAGA AGAATTATTT GAACGCTGTT TCCAAGAAAT GCTGGAAGAG
351 GAAGAAGAGC ATGAATGGTT TATCCAGCT CGAGATCTCC CACAACTAT
401 GGACCAAATC CAAGACCAGT TTAATGACCT TGTATCAGT GAAGGCTCTT
451 CTCTGGAAGA TCTTGTGGTC AAGAGCAATC TGAATCCAAA TGCAAAGGAG
501 TTTGTTCTTG GGGTGAAGTA CGGAAATATT TGAGTAGACG GGGCCCTCTT
551 TTGGTGGATG TAGCACAAAT TCCCACTGT GAAGGCAGTA TTAGAAGACT
601 TAATTGTAAA AGCACTCTTG TCACTGTGTT AACTTATGC ATTGCCAAAG
651 TTTTGTGTAG TCTTGCATGC TTAATAAAAG TGCTGAGACT GTTACTAAGT
701 AAAAAGCTGT CAAACATTTA CTGAAAATAG AATTGGCCCC ATGGCTTGAT
751 GTGAAGACAG CAAGGAAAGA AGCACCAGTC AAGTTGTGAA CAAGCACCAA
801 ATTAAGAGCT CTAACCTTA CCAAATTGTC TTTTTTTGAG GCTAATCTAT
851 CACTTGTAA TGTCTAACT TTAATATCAG TACATTAAAT TTGAGTTCCA
901 ACTGTTAAGC ATATTTCTCA GACTTAAATT TGATTATGTC CCCATCAAAA
951 AGAATCTCCA TTTTCTGAAG GTCTGTTAGT TAATTGAGA TAATTGTGA
1001 AAGGCAAGTA TGTCAATATA CTGAGGCTAC AAGTTAGTCA GCAGATGAGT
1051 GCCAGTCCAG CCTTTCCGG TATGTTATTG TTAGAAATAT TGAGTTCTAA
1101 TGTACATCT GAGGAAGTAT GTAATTTGAG AATTGTAAT TCTAAGGGAT
1151 TCACTGCATC ATAGCTATGC CTGTATGGAG TCTAACATAT GACCAATACC
1201 AACCATAAAT CCAGCTGAAC AAAGATACTG TAACATTATG ATTTGAGTGG
1251 TGCTTTTCTT TGCTTTGTGA ACCATCACGA GAGTCTGCAG CACAACCTTT
1301 AACAAAGCTA GAACAGTTT GGCTTCTTAA ACTTCATATT TGGGTAGGTT
1351 AAGCTGCCAT ACGTGTTTCA TGTGAATAGT GTTTAAGTTG AAAATATTGT
1401 AAAAAAATTA TATTTTTC AAAAAATTTA AAAAAATAAA TAATAGTAGA
1451 ACTGAAAAAA AAAAAA AAAAATAAAA AAAAAATAAA AAAAAATAAA
1501 AAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 150 bp to 530 bp; peptide length: 127
Category: putative protein

1 MKDPSRSSTS PSIINEDVII NGHSHEDDNP FAEYMMWENE EEFNRQIEEE

51 LWEEEFIERC FQEMLEEEEE HEWFIPARDL PQTMDQIQDQ FNDLVISEGS
101 SLEDLVVKS NLPNAKEFVP GVKYGNI

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3f16, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_3f16, frame 3

Report for DKFZphfbr2_3f16.3

[LENGTH] 127
[MW] 14998.41
[pI] 4.04
[BLOCKS] BL01269D
[PROSITE] MYRISTYL 1
[PROSITE] CK2_PHOSPHO_SITE 2
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 27.56 %

SEQ MKDPSRSSTSPSIINEDVIINGHSHEDDNPFAEYMWENEEFNRIEEELWEEEFIERC
SEGXXXXXXXXXXXXXXXXXXXX
PRD ccc

SEQ FQEMLEEEEEHEWFIPARDLPQTMDQIQDQFNDLVISEGSLEDLVVKS NLPNAKEFVP
SEGXXXXXXXXXXXX
PRD hhhhhhhhhhhhhcc

SEQ GVKYGNI
SEG
PRD ccccccc

Prosite for DKFZphfbr2_3f16.3

PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	100->104	CK2_PHOSPHO_SITE	PDOC00006
PS00008	121->127	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_3f16.3)

DKF2phfbr2_3g8

group: metabolism

DKF2phfbr2_3g8.1 encodes a novel 178 amino acid protein with similarity to yeast ARD1 protein.

In yeast, ARD1 and NAT1, are required for the expression of an N-terminal protein acetyltransferase 1. NAT1 controls full repression of the silent mating type locus HML, sporulation and entry into G0. ARD1 is involved in the assembly of the NAT 1-complex. The new protein could be part of this or an other NAT complex.

The new protein can find application modulating NAT assembly and action and therefore be important in metabolism of drugs and environmental mutagens.

strong similarity to N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 homolog

complete cDNA, complete cds? start at Bp 40, EST hits

Sequenced by AGOWA

Locus: /map="20"

Insert length: 1030 bp

Poly A stretch at pos. 1013, no polyadenylation signal found

```

1 TGGGCTTGGC GAACGGTCTT CGGAAGCGGC GCGGCGCGA TGACCACGCT
51 ACGGGCCTTT ACCTGCGACG ACCTGTTCCG CTTCAACAAC ATTAACCTGG
101 ATCCACCTTAC AGAAACTTAT GGGATTCCTT TCTACCTACA ATACCTCGCC
151 CACTGGCCAG AGTATTTTCAT TGTTCAGTG GCACCTGGTG GAGAATTAAT
201 GGGTTATATT ATGGGTAAAG CAGAAGGCTC AGTAGCTAGG GAAGATGGC
251 ACGGGCACGT CACAGCTCTG TCTGTTGCCC CAGAATTTCG ACGCTTGGT
301 TTGGCTGCTA AACTTATGGA GTTACTAGAG GAGATTTTCAG AAAGAAAGGG
351 TGGGTTTTTT GTGGATCTCT TTGTAAGAGT ATCTAACCAA GTTGCACTTA
401 ACATGTACAA GCAGTTGGGC TACAGTGTAT ATAGGACGGT CATAGAGTAC
451 TATTCGGCCA GCAACGGGGA GCCTGATGAG GACGCTTATG ATATGAGGAA
501 AGCACTTTCC AGGGATACTG AGAAGAAATC CATCATACCA TTACCTCATC
551 CTGTGAGGCC TGAAGACATT GAATAACCCT GGGCAGTGGT TCTTAGGCAG
601 ATACTCTAGA TGCTTTATGG ACAATATTAT TTTCATTGGA TGATTCTGGA
651 GCTCTATTAG GAGAAAAGTA ATCATTTTAG GTCTTAAAGA CTTCAAGAAA
701 ATACAGGTTA TCAATTTATT TTAATCTCTA TTGTTTCCAG TTAGCAATAT
751 CATACCTATT AAAGCTGTTC ATTGTAACAA AATTCAATCA AAAAGGCAGC
801 TAGGTCAGAA GGAACATAC CACTCTCATG GTTCATAGTA TTCCTGTAT
851 GTATGCTAGG GAAAGACTT GCTCCAGTCT CCTCCTCAGT TCTGTGCCTG
901 AGAACCACCTG CTGCATATAT TTGTTTAA ATTTGTATT GAACGTGTTAA
951 TTGAAGCTTT AAAAGCATAT ATGAAATGTA TAAATCTAAG ATGTATAATA
1001 CATTATTGAC TCCAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HSG0101 from database EMBL:
human STS SHGC-35956.
Length = 401
Minus Strand HSPs:
Score = 1417 (212.6 bits), Expect = 9.3e-58, P = 9.3e-58
Identities = 301/311 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 40 bp to 573 bp; peptide length: 178
Category: strong similarity to known protein

```

1 MTLRAFTCD DLFRFNNINL DPLTETYGIP FYLQYLAHP EYFIVAVAPG
51 GELMGYIMGK AEGSVAREEW HGHVTALSVA PEFRRGLAA KLMELLEIS

```


101 ERKGGFFVDL FVRVSNQVAV NMYKQLGYSV YRTVIEYISA SNGEPDEDAY
151 DMRKALSRTD EKSIIPPLPH PVRPEDIE

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_3g8, frame 1

TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4., N = 1, Score = 475, P = 3.2e-45

SWISSPROT:ARDH LEIDO N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG., N = 1, Score = 451, P = 1.1e-42

PIR:S69021 hypothetical protein YPR131c - yeast (Saccharomyces cerevisiae), N = 1, Score = 382, P = 2.3e-35

>TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. Length = 180

HSPs:

Score = 475 (71.3 bits), Expect = 3.2e-45, P = 3.2e-45
Identities = 96/165 (58%), Positives = 118/165 (71%)

Query: 1 MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGE--LMGYIM 58
MT R F DLF FNNINLDPLTET+ I FYL YL WP +V + + LMGYIM
Sbjct: 1 MTDTRKFKATDLFSFNNINLDPLTETFNISFYLSYLNKWPSCVQESDLSDPTLMGYIM 60

Query: 59 GKAEGSVAREEWHGHVLTALSVAPFRRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQV 118
GK+EG+ +EWH HVTA++VAP RRLGLA +M+ LE + + FFVDLFVR SN +
Sbjct: 61 GKSEGT--GKEWHTHTVTAITVAPNSRRLGLARTMDYLETVGNSENAFFVDLFVRASNAL 118

Query: 119 AVNMYKQLGYSVYRTVIEYISASNGEPDEDAYDMRKALSRTDEKSI 165
A++ YK LGYSVYR VI YYS +G+ DED++DMRK LSRD ++SI
Sbjct: 119 AIDFYKGLGYSVYRRVIGYISNPHGK-DEDSFDMRKPLSRDVRNRESI 164

Pedant information for DKFZphfbr2_3g8, frame 1

Report for DKFZphfbr2_3g8.1

[LENGTH] 178
[MW] 20338.24
[pI] 5.06
[HOMOL] TREMBL:SPCC16C4_12 gene: "SPCC16C4.12"; product: "putative n-terminal acetyltransferase complex subunit"; S.pombe chromosome III cosmid c16C4. 7e-47
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YPR131c] 6e-37
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YHR013c] 4e-14
[FUNCAT] r general function prediction [M. jannaschii, MJ1530] 6e-09
[PIRKW] acyltransferase 1e-12
[SUPFAM] arrest-defective protein 1 1e-12
[SUPFAM] Escherichia coli peptide N-acetyltransferase rimI 1e-07
[PROSITE] CK2_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 3
[KW] Alpha_Beta

SEQ MTTLRAFTCDDLFRFNNINLDPLTETYGIPFYLAHWPEYFIVAVAPGGELMGYIMGK
PRD cccccccccchhhhhccccccccccccchhhhhccccceeeeeeccccceehhhh

SEQ AEGSVAREEWHGHVLTALSVAPFRRRLGLAAKLMELLEISERKGGFFVDLFVRVSNQVAV
PRD hccccccccccccceeeehhhhhhhccchhhhhhhhhhhhhccccceeeeeeccccchhhh

SEQ NMYKQLGYSVYRTVIEYISASNGEPDEDAYDMRKALSRTDEKSIIPPLPHVVRPEDIE
PRD hhhhhccccchhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccc

Prosite for DKFZphfbr2_3g8.1

WO 01/12659

PCT/IB00/01496

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	100->103	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	8->12	CK2_PHOSPHO_SITE	PDOC00006
PS00006	133->137	CK2_PHOSPHO_SITE	PDOC00006
PS00006	141->145	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKF2phfbr2_3g8.1)

DKFZphfbr2_312

group: brain derived

DKFZphfbr2_312 encodes a novel 589 amino acid protein with weak similarity to *S. cerevisiae* ubiquitin-like protein DSK2.

Pfam predicts for this protein similarity to the ubiquitin family; No informative BLAST results; No predictive prosite or SCOP motive

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ubiquitin-like protein DSK2 yeast

complete cDNA, complete cds, EST hits
Dsk2p is involved in spindle pole body SPB duplication, SPB = centomer
strong similarity to HRIHFB2157 human mRNA

Sequenced by AGOWA

Locus: unknown

Insert length: 2978 bp
Poly A stretch at pos. 2958, polyadenylation signal at pos. 2924

```

1  GGGGGGAGGA AGCGGTGGCT GCTGCGGATG TCGGTGTGAG CGAGCGGGCGC
51  CTGAACACAC GGCGGCTGCC GAGCGCCTGA CCCGGGCTCG CGCCAGAGCC
101  TGCACCGAGC TCCGGGGCCC CACACCCGCT ACGGTGGCCC TCGCCCGCTT
151  GCTACTGAGG CGGCGTGCTC TGCATTCTTC GCTGTCCAGG CCTGCCGGCT
201  CTGGTGCTTG CTGGCTCCTC CTTGCTCGCC TGCTCCCTCC TGCTTGCCCTG
251  AGTCACCGCC GCGCGCGCCG CCACAGCCAT GGCCGAGAGT GGTGAAAGCG
301  GCGGTCTCTC GGGCTCCCAG GATAGCGCCG CCGGAGCCGA AGGTGCTGGC
351  GCCCCGCGGG CCGCTGCCCTC CGCGGAGCCC AAAATCATGA AAGTCACCGT
401  AAGAGCCCCG AAGGAAAAGG AGGAATTCGC CGTGCCCGAG AATAGCTCCG
451  TCCAGCAGTT TAAGGAAGAA ATCTCTAAAC GTTTTAAATC ACATACTGAC
501  CAACTTGTGT TGATATTTGC TGGAAAAATT TTGAAAGATC AAGATACCTT
551  GAGTCAGCAT GGAATTCATG ATGGACTTAC TGTTCACCTT GTCATTAATA
601  CACAAACACG GCCTCAGGAT CATTACGCTC AGCAAAACAA TACAGCTGGA
651  GGCATGTGTA CTACATCATC AACTCCTAAT AGTAACTCTA CATCTGGTTC
701  TGCTACTAGC AACCCTTTTG GTTTAGGTGG CCTTGGGGGA CTGCAAGTTC
751  TGAGTAGCTT GGGTTTGAAT ACTACCAACT TCTCTGAATC ACAGAGTCAG
801  ATGCAGCGAC AACTTTTGTC TAACCCGTAA ATGATGGTCC AGATCATGGA
851  AAATCCCTTT GTTCAGAGCA TGCTCTCAAA TCCTGACCTG ATGAGACAGT
901  TAATTATGGC CAATCCACAA ATGCAGCAGT TGATACAGAG AAATCCAGAA
951  ATTAGTCATA TGTGAAATAA TCCAGATATA ATGAGACAAA CGTTGGAACT
1001  TGCCAGGAAT CCAGCAATGA TGCAGGAGAT GATGAGGAAC CAGGACCGAG
1051  CTTTGAGCAA CCTAGAAAGC ATCCCAGGGG GATATAATGC TTTAAGGGCGC
1101  ATGTACACAG ATATTCAGGA ACCAATGCTG AGTGCTGCAC AAGAGCAGTT
1151  TGGTGGTAAT CCATTGCTT CCTTGGTGAG CAATACATCC TCTGGTGAAG
1201  GTAGTCAACC TTCCCGTACA GAAAATAGAG ATCCACTACC CAATCCATGG
1251  GCTCCACAGA CTTCCCAGAG TTATCAGCTC TCCAGCGGCA CTGCCAGCAC
1301  TGTGGGTGGC ACTACTGGTA GTACTGCCAG TGGCACTTCT GGGCAGAGTA
1351  CTACTGCGCC AAATTTGGTG CCTGGAGTAG GAGCTAGTAT GTTCAACACA
1401  CCAGGAATGC AGAGCTTGTT GCAACAAATA ACTGAAACC CACAACATGAT
1451  GCAAAACATG TTGTCTGCCC CCTACATGAG AAGCATGATG CAGTCACTAA
1501  GCCAGAATCC TGACCTTGCT GCACAGATGA TGCTGAATAA TCCCCTATTT
1551  GCTGGAATTC CTCAGCTTCA AGAACAAATG AGACAAACAG TCCCAACTTT
1601  CCTCCAACAA ATGCAGAAAT CTGATACACT ATCAGCAATG TCAAAACCTA
1651  GAGCAATGCA GGCCTTGTTA CAGATTACAG AGGGTTTACA GACATTAGCA
1701  ACGGAAGCCC CGGGCCTCAT CCCAGGGTTT ACTCCTGGCT TGGGGGCATT
1751  AGGAAGCACT GGAGGCTCTT CGGGAACATA TGGATCTAAC GCCACACCTA
1801  GTGAAACAC AAGTCCCACA GCAGGAACCA CTGAACCTGG ACATCAGCAG
1851  TTTATTCAGC AGATGCTGCA GGCTCTTGCT GGAGTAAATC CTCAGCTACA
1901  GAATCCAGAA GTCAGATTTC AGCAACAACT GGAACAACTC AGTGCAATGG
1951  GATTTTTGAA CCGTGAAGCA AACTTGCAAG CTCTAATAGC AACAGGAGGT
2001  GATATCAATG CAGCTATTGA AAGGTTACTG GGCTCCAGC CATCATAGCA
2051  GCATTTCTGT ATCTTGAAAA AATGTAATTT ATTTTGTATA ACGGCTCTTA
2101  AACTTTAAAA TACCTGCTTT ATTTTATTTT GACTCTTGGA ATTTCTGTGT
2151  GTTATAAACA AACCAATAT GATGCATTTT AAGGTGGAGT ACAGTAAGAT
2201  GTGTGGGTTT TTCTGTATTT TTCTTTTCTG GAACAGTGGG AATTAAGGCT
2251  ACTGCATGCA TCACTTCTGC ATTTATTGTA ATTTTAAAA AACATCACCT
2301  TTTATAGTTG GGTGACCAGA TTTTGTCCCT CATCTGTCCA GTTTATTTGC
2351  TTTTAAACA TTAGCCTATG GTAGTAATTT ATGTAGAATA AAAGCATTAA
2401  AAAGAAGCAA ATCATTTGCA CTCTATAATT TGTGGTACAG TATTGCTTAT
2451  TGTGACTTTG GCATGCATTT TTGCAACAAA TGCTGTAAGA TTTATACTAC
2501  TGATAATTTT GTTTTATTTG TATACAATAT AGAGTATGCA CATTGGGGAC
```

```

2551 TGCATTTCTG GAAACATACT GCAATAGGCT CTCTGAGCAA AACACCTGTA
2601 ACTAAAAAAG TGAAGATAAG AAAATACTCT TAAAGCTGAG TATTTCTTAA
2651 TTGTATAGAA TCTTACAGCA TCTTTGACAA ACATCTCCCA GCAAAAGTGC
2701 CGGTTAGTCA GGTTTGTGTA AAATACAGTA GAAAAGCTGA TTCTGGTTAT
2751 CTCTTTAAGG ACAATTAATT GTACAGACAC ATAATGTAAC ATTGTCTCAA
2801 CATTTCATCA CAGATTGACT GTAAATTACC TTAATCTTTG TGCAGACTGA
2851 AGGAACACTG TAGTATACCC CAAAGTGCAAT TTGCCTAGGA CTTCTCAGCT
2901 TCTCCCATAG GTAGTTTAAC AGGCATTAAA ATTTGTAATT GAAATGTTGC
2951 TTTCACCTCA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 279 bp to 2045 bp; peptide length: 589
 Category: similarity to known protein

```

1 MAESGESGGP PGSQDSAAGA EGAGAPAAAA SAEPKIMKVT VKTPKEKEEF
51 AVPENSSVQQ FKEEISKRFK SHTDQLVLIF AGKILKDQDT LSQHGIDHGL
101 TVHLVIKTN RPQDHSAAQT NTAGGNVTS STPNNSSTSG SATSNPFGLG
151 GLGGLAGLSS LGLNTTNFSE LQSQMRQLL SNPEMMVQIM ENPFVQSMLS
201 NPDLMRQLIM ANPQMQLIQ RNPEISHMLN NPDIMRQTL LARNPAMMQE
251 MMRNQDRALS NLESIPGGYN ALRRMYTDIQ EPMLSAEQE FGGNPFASLV
301 SNTSSGESGQ PSRTENRDPL PNPWAPQTSQ SSSASSGTAS TVGGTTGSTA
351 SGTSSGQSTTA PNLVPGVGAS MFNTPGMQSL LQQITENPQL MQNMLSAPYM
401 RSMMQSLSQN PDLAAQMLN NPLFAGNPQL QEQRQLPT FLQQMQNPDT
451 LSAMSNPRAM QALLIQQGL QTLATEAPGL IPGFTPLGA LGSTGGSSGT
501 NGSNATPSN TSPTAGTTEP GHQQFIQQL QALAGVNPQL QNPEVRFQQQ
551 LEQLSAMGFL NREANLQALI ATGGDINAAI ERLGSSQPS

```

BLASTP hits

Entry CE1_1 from database TREMBL:

"F15C11.2"; Caenorhabditis elegans cosmid VF15C11L

Length = 293

Score = 454 (159.8 bits), Expect = 4.4e-43, P = 4.4e-43

Identities = 81/162 (50%), Positives = 113/162 (69%)

Entry S54583 from database PIR:

ubiquitin-like protein DSK2 - yeast (Saccharomyces cerevisiae)

Length = 373

Score = 278 (97.9 bits), Expect = 1.2e-23, P = 1.2e-23

Identities = 100/307 (32%), Positives = 155/307 (50%)

Entry AB015344.1 from database TREMBLNEW:

gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial cds.

Score = 1135, P = 3.6e-115, identities = 227/301, positives = 253/301

Alert BLASTP hits for DKFZphfbr2_312, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_312, frame 3

Report for DKFZphfbr2_312.3

```

[LENGTH]      589
[MW]           62489.22
[pI]           5.02
[HOMOL]        TREMBL:AB015344.1 gene: "HRIHFB2157"; Homo sapiens HRIHFB2157 mRNA, partial
cds. 1e-121
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YMR276w] 2e-17

```

```

(FUNCAT)      30.10 nuclear organization      [S. cerevisiae, YMR276w] 2e-17
[BLOCKS]      BL00299 Ubiquitin family proteins
[SUPFAM]      unassigned ubiquitin-related proteins 5e-16
[SUPFAM]      ubiquitin homology 5e-16
[PROSITE]     MYRISTYL 24
[PROSITE]     CK2_PHOSPHO_SITE 9
[PROSITE]     GLYCOSAMINOGLYCAN 1
[PROSITE]     PKC_PHOSPHO_SITE 3
[PROSITE]     ASN_GLYCOSYLATION 7
[PFAM]        Ubiquitin family
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY 23.43 %

```

```

SEQ      MAESGESGGPPGSQDSAAGAEGAGAPAAAAAEPKIMKVTVKTPKEKEEFVFPENSSVQQ
SEG      ..xxxxxxxxxxxx..xxxxxxxxxxxxxxxxxxxx..xxxxxxxxxxxx..
laarA    .....CEEEEEETTCEEEECTTTTBHHH

SEQ      FKEEISKRFKSHTDQLVLIFAGKILKQDQLSQHGIDGLTVHLVIKTNRPQDHSAAQT
SEG      .....
laarA    HHHHHHHHHCCCGGEEEEETTECTTTTBGGGGCCTTTTEEEEBEBC.....

SEQ      NTAGGNVTSTSPNSNSTSGSATSNPFGLGGLGGLAGLSSSLGNTNFSSELQSQMQRQLL
SEG      ..xxxxxxxxxxxxxxxxxxxxxxxx..xxxxxxxxxxxxxxxx..
laarA    .....

SEQ      SNPEMMVQIMENPFVQSMLSNPDLMRQLIMANPQMQLIQRNPEISHMLNPNPDIMRQTLE
SEG      .....
laarA    .....

SEQ      LARNPAMQEMMRNQDRALSNLESI PGGYNALRRMYTDIQEPMLSAAQEQFGGNPFASLV
SEG      .....
laarA    .....

SEQ      SNTSSGEGSQPSRTENRDLPLNPWPAPQTSQSSSASSTASTVGGTTGSGTSGQSTTA
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
laarA    .....

SEQ      PNLVPGVGASMFNTPGMQSLQITENPQLMQNMLSAPYMRSMMSLSQNPDLAAQMMLN
SEG      .....
laarA    .....

SEQ      NPLFAGNPQLQEQRQQLPTFLQMQNPDTLSAMSNPRAMQALLQIQOGLQTLATEAPGL
SEG      .....
laarA    .....

SEQ      IPGFTPLGALGSTGGSSGTNGSNATPSENTSPTAGTTEPGHQFQIQMLQALAGVNPQL
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxx
laarA    .....

SEQ      QNPEVRFQQQLEQLSAMGFLNREANLQALATGGDINAAIERLLGSQPS
SEG      .....
laarA    .....

```

Prosites for DKF2phfbr2_312.3

PS00001	55->59	ASN_GLYCOSYLATION	PDOC00001
PS00001	126->130	ASN_GLYCOSYLATION	PDOC00001
PS00001	136->140	ASN_GLYCOSYLATION	PDOC00001
PS00001	164->168	ASN_GLYCOSYLATION	PDOC00001
PS00001	167->171	ASN_GLYCOSYLATION	PDOC00001
PS00001	302->306	ASN_GLYCOSYLATION	PDOC00001
PS00001	501->505	ASN_GLYCOSYLATION	PDOC00001
PS00002	305->309	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	40->43	PKC_PHOSPHO_SITE	PDOC00005
PS00005	43->46	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	71->75	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	200->204	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	572->576	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	12->18	MYRISTYL	PDOC00008

PS00008	19->25	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	95->101	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	140->146	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	162->168	MYRISTYL	PDOC00008
PS00008	267->273	MYRISTYL	PDOC00008
PS00008	293->299	MYRISTYL	PDOC00008
PS00008	308->314	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	355->361	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	492->498	MYRISTYL	PDOC00008
PS00008	495->501	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	573->579	MYRISTYL	PDOC00008

Pfam for DKF2phfbr2_312.3

HMM_NAME	Ubiquitin family		
HMM	*MQIFVKTLtGRTcTFEVepQEtVeqIKQHieekEGIPPeQQRLLIFaGRQ		
	M ++VKT	+ +F V+++ V Q+K+ I+	+Q +LIFAG+
Query	37	MKVTVKTPK-EKEEFAVPENSSVQQFKEEISKRFKSHDQLVLIFAGKI	84
HMM	LEDeKTLsDYNiggeSTLHLVLR*		
	L D	TLS+++I + T+HLV++	
Query	85	LKDQDTLSQHGIDGLTVHLVIK	107

DKFZphfbr2_62b11

group: signal transduction

DKFZphfbr2_62b11 encodes a novel 655 amino acid putative GTPase-activating protein, related to human chimaerins.

The rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to CHIMAERIN

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="4"

Insert length: 4593 bp

Poly A stretch at pos. 4571, polyadenylation signal at pos. 4553

```
1 GGGGGAGTTT GAAGACAGAA AGGAAAGGGG AGAAACCTGC AGAGAGCATC
51 AAAGGATGGG GGGTGCTATA AAAGAAGCAG GGGGGTCCTT TGAAAGAAAT
101 CTATCATGCA CTGAAATGCT TTCTGGAGAA GGTGCCGTTA TTTTCTCCCC
151 CTCTTGCTCA GATGAAAGGA GCCAGCAAGG ACAGTCCTGA AATATTCCCTC
201 AGGGGACTTT TTGTCATTGT TCCTCTTTCC TCTTGCACAG AGCTATTTCG
251 TGACCTTTCC AGAGGAATCT CAGTCCAGCT GAGAAGACAG TTCTTAATAA
301 AAACAAAAAA ATGCAAAAAC CAATTCCTGC TGTTTGAATG GGAATGGTAG
351 CTTGCTTGCT GCAGTTCTTT TCCTGTGACA TTTTGGAATG TCTGCAGAAA
401 CTTAAAAAAA AGAAAAAAA AACCTTAAAA ACTCCCTGGA TTAGGCAAGA
451 GAAAGGAAG TTTTTTTTGT CTAACACGGA GTAAATGAGA GGTGGTAACT
501 TATCCCTAAG CCAGGACCTG GATGATCAAA ACCTTCAAAT TCTAGGGATC
551 AGCACTTCAA AAATAACAAG TAAACAAGCA TGAGGAGTGG CTGTTGGGTT
601 TCGCTCAGAG GCAGGTTTTA AAGGAAGCCA AAACCGGGTT CAGAACTTCA
651 GGCCTGTACG ATGCCTGAAG ACCGGAATTC TGGGGGGTGC CCGGCTGGTG
701 CCTTAGCCTC AACTCCTTTC ATCCCTAAAA CTACATACAG AAGAATCAAA
751 CGGTGTTTTA GTTTTCGGAA AGGCATTTTT GGACAGAAAC TGGAGGATAC
801 TGTTTCGTTT GAGAAGAGAT ATGGGAACCG TCTGGCTCCG ATGTTGGTGG
851 AGCAGTGCCT GGACTTTATC CGACAAAGGG GGCTGAAAGA AGAGGGTCTC
901 TTTGCACTGC CAGGCCAGGC TAATCTTGTT AAGGAGCTCC AAGATGCCCTT
951 TGACTGTGGG GAGAAGCCAT CATTTGACAG CAACACAGAT GTACACACGG
1001 TGGCATCACT TCTTAAGCTG TACCTCCGAG AACTTCCAGA ACCAGTTATT
1051 CCTATGCGA AGTATGAAGA TTTTTTGTCA TGTGCCAAAC TGCTCAGCAA
1101 GGAAGAGGAA GCAGGTGTTA AGGAATTAGC AAAGCAGGTG AAGAGTTTGC
1151 CAGTGGTAAA TTACAACCTC CTCAAGTATA TTTGCAGATT CTTGGATGAA
1201 GTACAGTCCT ACTCGGGAGT TAACAAAATG AGTGTGCAGA ACTTGGCAAC
1251 GGTCTTTGGT CCTAATATCC TGCGCCCAA AGTGGAAAGT CTTTGACTA
1301 TCATGGAGGG CACTGTGGTG GTCCAGCAGT TGATGTCAGT GATGATTAGC
1351 AAACATGATT GCCTCTTTCC CAAAGATGCA GAACTACAAA GCAAGCCCCA
1401 AGATGGAGTG AGCAACAACA ATGAAATTCA GAAGAAAAGCC ACCATGGGGC
1451 TGTTACAGAA CAAGGAGAAC AATAACACCA AGGACAGCCC TAGTAGGCAG
1501 TGCTCTGGG ACAAGTCTGA GTCACCCAG AGAAGCAGCA TGAACAATGG
1551 ATCCCCCACA GCTCTATCAG GCAGCAAAAC CAACAGCCCA AAGAACAGTG
1601 TTCACAAGCT AGATGTGTCT AGAAGCCCCC CTCTCATGGT CAAAAAGAAC
1651 CCAGCCTTTA ATAAGGGTAG TGGGATAGTT ACCAATGGGT CCTTCAGCAG
1701 CAGTAATGCA GAAGGTCTTG AGAAAACCCA AACCACCCCC AATGGGAGCC
1751 TACAGGCCAG AAGGAGCTCT TCACTGAAGG TATCTGGTAC CAAAATGGGC
1801 ACGCACAGTG TACAGAATGG AACGGTGGC ATGGGCATTT TGAACACGCA
1851 CACACTCGGG AACCCACAAA ATGTTGGAAG CATGAGCTGG CTGCCAAATG
1901 GCTATGTGAC CCTGAGGGAT AACAAAGCAG AAGAACAAGC TGGAGAGTTA
1951 GGCAGCACA ACAGACTGTC CACCTATGAT AATGTCCATC AACAGTTCTC
2001 CATGATGAAC CTTGATGACA AGCAGAGCAT TGACAGTGCT ACCTGGTCCA
2051 CTTCCTCTG TGAATCTCC CTCCCTGAGA ACTCCAATC CTGTGCTCT
2101 TCTACCAACA CCTGCCCAGA GCAAGACTTT TTTGGGGGGA ACTTTGAGGA
2151 CCCGTGTTTG GATGGGCCCC CGCAGGACGA CCTTCCACAC CCCAGGGACT
2201 ATGAAAGCAA AAGTGACCAC AGGAGTGTGG GAGGTGCAAG TAGTCGTGCC
2251 ACCAGTAGCA GTGACAACAG TGAGACATTT GTGGGCAACA GCAGCAGCAA
2301 CCCAGTGCA CTGCACAGTT TAGTTTCCAG CCTGAAACAG GAAATGACCA
2351 AACAGAAGAT AGAGATGAG TCCAGGATAA AGAGCTTAGA ACAGCGAAAC
2401 TTGACTTTGG AAACAGAAAT GATGAGCCTC CATGATGAAC TGGATCAGGA
2451 GAGGAAAAAG TTCACAATGA TAGAAATAAA AATGCGAAAT GCCGAGCGAG
2501 CAAAAGAGA TGCCGAGAAA AGAAATGACA TGCTACAGAA AGAAATGGAG
2551 CAGTTTTTTT CCACGTTTGG AGAACTGACA GTGGAACCCA GGAGAACCGA
```

```

2601 GAGAGGAAAC ACAATATGGA TTCAGTGAGC CTGCTTTTCGC CTGCTGTCTC
2651 TGATGGCTCT GGCAAGGACT CCAGGGGATTC TGGTGGGATA TGAAGGATAA
2701 CCAGGTGGCT GGTCACTGG ATGTACAGAA GTCTAACTGG TGAAGGATAA
2751 TCATTTACAG ACATTAAACA TCCATATCTG CAATGTGTAC CAAAGTTATA
2801 TCATGCCCA TAATGCTACT GTCAAGTGT AACTGGAT ATGTGTATAT
2851 AGAGTAGTTT TTCAAAAGTA AACTAAAAAT GAGAAGCATA TTCAAGAAAT
2901 TATTTTATTG CAAGTCTTGT ATTTAAATGT TAAATCAATA TGTTGTTGCA
2951 ATTTAGCTTG CTTTCAAGCT TCACCCCTTG CACTTAACAT AAGCTATTTT
3001 TGGCATTGTG TTATCATCGG CTTATTTTAT AGATCAATAT TTTTATTTC
3051 CTTTTTGTCT GAGGAAATGA AGATAAGCAA AAATATAAAT ATATATATAA
3101 ATATATGAGT TATTAAACC AGAAGAATAC TTTGTGGCTG TGCTGTTTGT
3151 GCCAATAGAC TTTGTCATGA CCAAAAAGAG AAATGTAAT AGTTTATATA
3201 AATACAGTCG AATCACCAGG AACCTTTGAG CTGCTTTTAA AATTCTTCCC
3251 CTGGCACCAC TCAGTTTTCG TTTTGGCAGG CGATTTGACA TAGGAACCTT
3301 GAGACTCCAG GAGAAAGTCC CTTTCTGAGG CCCACTGTCT ACCTTGCCAG
3351 ATCCTCAGTG CGTATCGCCA ATGCAGGATG CTCCTTAGAA AAGAAAAAAT
3401 GGTAAGGATG GGCATTTAAC GATTCAGGCT TTGAATTACT CTGTCCTCTC
3451 GGACCGAATC TCTTTAACTG CTGGATAGTT TTAGAGGAAT TCTCCTGCTA
3501 CTTAGGTACT GGGAAACAAT GCTTGCTAAA CCATGCCAC GTGAGCACCT
3551 GTCTCCCCT CAAACCTCTC CCATCTCCCA ACAACTGCAC TTAGAATAC
3601 CAGCAGTGAA ATGGTATTAC TGTTTCCCTC TGAGTGAAC TGCTAGAGTA
3651 TATGTCACGT AGTGACATT TTTTCTCACT CAGGCTATTG CCATCTGGGA
3701 TTCTCTCCCT ACTACAGCTG GCAAAGTTGG TTTGCAGCAA GAAGATAGTG
3751 GGAGGGGGCC AGGCTGCAGG AGAAGGAGAA AAGTTTAGAA GAAACAAACC
3801 ATTTTGCTTC TAATTTTGAC AGTATCACTT TCCTGTTAAA ACATACAATA
3851 ATTTTAAAG GTGAATGCCT AAAGTTCCAA TTTTAGCAAA TATGGGAACC
3901 TCAGCAATGC TAATTTTCTA GAAAACCCA GGGCTCTTTG GAGCTAGAGT
3951 TTTGGGAGAA CAGTTCTTCA CAATAAGGCA ATGGTTTGA GAGGCCAGGC
4001 AAATAATCTT TCTCACCGTA GAACAAAAG TTACAAAAGG CATAATCGGA
4051 AATAGAGACT ACATACTTGA GTTTATGGG TTTGTGTTGT TTGAAGGTTT
4101 AATGCTTGCA TGTGTTTATT TATTTTCAAG AGGGAAAGTG GTCTGTACTG
4151 CTTTCATCCT TGCCACTGTC TTGCTTTTAT TTTTACTCT CCCACTGAGC
4201 AAGCGTCTGT GGTCCATG GGTCCATG TATCAACCAG TATCTTTATA GCAATAATTT
4251 CTTTAATTC CTTTCTCTC TCTTCCAAT TATTTAACC GTTACTTCCA
4301 CCTGGACATA CGATAGGAAA TTCAAATCA AAATATGAAA ATTGATCTTA
4351 ATAACCTCTC CTTATATCT TTTCACCTAT TTCCAGTCT TATCATAGTT
4401 GATAAAACC TCAGACTCAT CCAGAAAGCT ATATGATGCA CTAGTAAAAA
4451 AAACAAAGAT ATTTAACTG CTGGGTTCA AATGGTATAC AATTTGCCAG
4501 CTGTTACTGA ACCTTCTATG CATAACTTTT TTTTCTCTCT GTGCAATTGG
4551 AATAATAAAA ATACTACTCC CATAAAAAA AAAAAAAA AAC

```

BLAST Results

Entry G38474 from database EMBLNEW:
 SHGC-58303 Human Homo sapiens STS genomic, sequence tagged site.
 Score = 2175, P = 1.2e-92, identities = 439/441

Medline entries

97476250:
 Beta2-chimaerin is a high affinity receptor for the phorbol ester tumor promoters.

Peptide information for frame 1

ORF from 661 bp to 2625 bp; peptide length: 655
 Category: similarity to known protein

```

1 MPEDRNSGGC PAGALASTPF IPKTTYRRIK RCFSEFRKIF GQKLEDTVRY
51 EKRYGNRLAP MLVEQCVDPI RQRLKEEGL FRLPGQANLV KELQDAFDCG
101 EKPSFDSNTD VHTVASLLKL YLRELPEPVI PYAKYEDFLS CAKLLSKEEE
151 AGVKELAKQV KSLPVVNYNL LKYICRFLDE VQSYSGVNKM SVQNLATVFG
201 PNILRPKVED PLTIMEGTVV VQQLMSVMIS KHDCLFPKDA ELQSKPQDGV
251 SNNNEIQKKA TMGLLQNKEN NNTKDSPSRQ CSWDKSESPO RSSMNNGSPT
301 ALSGSKTNSP KNSVHKLDVS RSPPLMVKKK PAFNKGSGIV TNGSFSSSNA
351 EGLEKTQTPP NGSQARRSS SLKVSQTRKM THSVQNGTVR MGILNSDTLG
401 NPTNVRNMSW LPNGYVTLRD NKQKEQAGEL GQHNRLSTYD NVHQQFSMMN
451 LDDKQSIDSA TWSTSSCEIS LPENSNSCRS STTTCPEQDF FGGNFEDPVL
501 DGPPQDDLSH PRDYESKSDH RSVGRSSRA TSSSDNSETF VGNSSSNHSA
551 LHSLVSSLKQ EMTKQKIEYE SRIKSLEQRN LTLETMMMSL HDELDQERKK

```


601 FTMIEIKMRN AERAKEDAEK RNDMLQKEME QFFSTFGELT VEPRRTERGN
651 TIWIQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62b11, frame 1

SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053., N = 3, Score = 661, P = 2.4e-89

TREMBL:HSU90908_1 product: "unknown"; Human clones 23549 and 23762 mRNA, complete cds., N = 1, Score = 348, P = 1.1e-29

PIR:S29128 N-chimerin - rat, N = 1, Score = 286, P = 2.8e-24

PIR:S29956 beta-chimerin - rat, N = 1, Score = 279, P = 1.6e-23

TREMBL:AB014572_1 gene: "KIAA0672"; product: "KIAA0672 protein"; Homo sapiens mRNA for KIAA0672 protein, complete cds., N = 1, Score = 314, P = 1e-24

>SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053.
Length = 638

HSPs:

Score = 661 (99.2 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 122/209 (58%), Positives = 160/209 (76%)

Query: 38 GIFGQKLEDTVRYEKRYGNRLAPMLVEQCVDIFIRQRLKEEGLFRLPGQANLVKELQDAF 97
G+FGQ+L++TV YE+++G L P+LVE+C +FI + G EEG+FRLPQG NLVK+L+DAF
Sbjct: 148 GVFGQRLDETVAIEQKFGPHLVPIVLEKCAEFIEHGRNEEGIFRLPGQDNLVKQLRDAF 207

Query: 98 DCGEKPSFDSNTDVHTVASLLKLYLRELPEPVIPIYAKYEDFLSCAKLLSKEEEAGVKELA 157
D GE+PSFD +TDVHTVASLLKLYLR+LPEPV+P+++YE FL C +L + +E +EL
Sbjct: 208 DAGERPSFDRDSDVHTVASLLKLYLRDLPEPVVPWSQYEGFLCGQLTNADEAKAQQELM 267

Query: 158 KQVKSLPVVNYNLLKYICRFLDEVQSYSGVNKMSVQNLATVFGPNILRPKVEDPLTIMEG 217
KQ+ LP NY+LL YICRFL E+Q VNKMSV NLATV G N++R KVEDP IM G
Sbjct: 268 KQLSILPRDNYSLSYICRFLHEIQLNCAVNKMSVDNLATVIGVNLIRSKVEDPAVIMRG 327

Query: 218 TVVVQQLMSVMISKHDCFLPKDAELQSKP 246
T +Q++M++MI H+ LFPK ++ P
Sbjct: 328 TPQIQRVMTMMIRDHEVLFPKSKDIPLSP 356

Score = 210 (31.5 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 45/115 (39%), Positives = 73/115 (63%)

Query: 531 TSSSDNSETFVGNSSSNHSLHSL---VSSLKQEMTKQKIEYESRIKSLEQRNLTLETM 587
T +S NSET G +S + SL V L++E+ QK YE +IK+LE+ N + ++
Sbjct: 523 TLASPNSETGPGKNSGEEIDSQRMVQELRKEIETQKQMYEEQIKNLEKENYDVWAKV 582

Query: 588 MSLHDELQERKKFTMIEIKMRNAERAKEDAEKRNDMLQKEMEQQFFSTFGELTVE 642
+ L++EL++E+KK +EI +RN ER++ED EKRN L++E+++F + E E
Sbjct: 583 VRLNEELEKEKKKSAALEISLRNMERSREDVEKRNKALEEEVKEFVKSMKEPKTE 637

Score = 70 (10.5 bits), Expect = 1.2e-74, Sum P(3) = 1.2e-74
Identities = 28/121 (23%), Positives = 54/121 (44%)

Query: 528 SRATSSSDNSETFVGNSSSNHSLHSLVSSLKQE-MTKQKIEYESRIKSLEQRNL-TLET 585
S+ TS+ DN + G+ SAL S K + + E K+ + + +L+
Sbjct: 489 SQRTSTYDNVPSLPGSPGEEASALSSQACDSKGDTLASPNSETGPGKNSGEEIDSQRL 548

Query: 586 EMMSLHDELQERKKFTMIEIKMRNAERAKEDAEKRNDMLQKEMEQQFFSTFGELTVEPRR 645
+ L E++ +++ M E +++N E+ D + L +E+E+ L + R
Sbjct: 549 MVQELRKEIETQKQ---MYEEQIKNLEKENYDVWAKVVRLNEELEKEKKKSAALEISLRN 605

Query: 646 TER 648
ER
Sbjct: 606 MER 608

Score = 53 (8.0 bits), Expect = 2.4e-89, Sum P(3) = 2.4e-89
Identities = 31/111 (27%), Positives = 46/111 (41%)

Query: 344 SFSSSNAEGLEKTQTPNGSLQARRSSSLKVSQTKMGTHSVQNG----TV---RMGILNSD 397
SFSS ++ + T T A S KV KG +Q+ T+ R L S
Sbjct: 388 SFSSMTSDS-DTTSPTGQQPSDAFPEDSSKVPREKPGDWKMQSRKRTQTLNPKCFLTSA 446

```

Query:   398 TLG-NPTNV---RMSWLPNGYVTLRDNKQKEQAGELGQ---HNRLSTYDNV 442
          G N + + +N W P + + + + +L Q R STYDNV
Sbjct:   447 EQGANSKMEIFKNEFWSPSSEAKAGEGHRRTMSQDLRQLSDSQTSTYDNV 498

Score = 53 (8.0 bits), Expect = 3.5e-14, Sum P(3) = 3.5e-14
Identities = 32/125 (25%), Positives = 56/125 (44%)

Query:   242 LQSKPDQD---VSNNIEQKKATMGLQNKEN--NNTKD---SPSRQCSWDKSESQRSS 293
          ++SK +D + +IQ+ TM ++++ E +KD SP Q + K RSS
Sbjct:   314 IRSKVEDPAVIMRGTPIQIRVMTM-MIRDHEVLFPKSKDIPLSPPAKQNDPKKAPVARSS 372

Query:   294 MNNGSPTALSGSKTNSPKNSVHKLDVRSRPLMVKKNPAFNKGSGIVTNGSFSSSNAEGL 353
          + + L S+T+S + D + P + + AF + S V +
Sbjct:   373 VGWDATEDLRISRTDSFSSMTSDSDTTS--PTGQQPSDAFPEDDSSKVPREKPGDWMQSR 430

Query:   354 EKTQTTPN 361
          ++TQT PN
Sbjct:   431 KRTOTLPN 438

```

Pedant information for DKFZphfbr2 62b11, frame 1

Report for DKFZphfbr2 62b11.1

```

[LENGTH] 655
[MW] 73394.60
[pI] 8.13
[HOMOL] SWISSPROT:Y053_HUMAN HYPOTHETICAL PROTEIN KIAA0053. 3e-71
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
          [S. cerevisiae, YPL115c] 1e-16
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 1e-16
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YPL115c]
1e-16
[FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 1e-16
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-16
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-16
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YDR379w] 4e-16
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-15
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 2e-13
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 2e-13
[SCOP] dlrgp_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens) 2e-46
[SCOP] dlpbwa_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Hom 6e-37
[PIRKW] phosphotransferase 3e-13
[PIRKW] breakpoint cluster region 2e-20
[PIRKW] transmembrane protein 7e-14
[PIRKW] brain 2e-20
[PIRKW] alternative splicing 2e-20
[PIRKW] P-loop 9e-19
[PIRKW] cytoskeleton 1e-08
[SUPFAM] CDC24 homology 7e-21
[SUPFAM] bcr protein 7e-21
[SUPFAM] myosin motor domain homology 9e-19
[SUPFAM] pleckstrin repeat homology 2e-15
[SUPFAM] LIM metal-binding repeat homology 9e-15
[SUPFAM] protein kinase C zinc-binding repeat homology 5e-24
[PROSITE] MYRISTYL 16
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 15
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 11
[PROSITE] ASN_GLYCOSYLATION 8
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 6.87 %
[KW] COILED_COIL 12.06 %

```

SEQ	MPEDRNSGGCPAGALASTPFIPKTTYRRIKRCFSFRKGI FGQKLEDTVRYEKRYGNRLAP
SEG
COILS
lr _g p-C
SEQ	MLVEQCVD FIRQRLKEEGLFRLPGQANLVKELQDAFDCGEKPSFDSNTDVHTVASLLKL
SEG
COILS
lr _g p-	HHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHCCCCGGGCCCHHHHHHHHH
SEQ	YLRELPEVPVIPYAKYEDFLSCAKLLSKEEEAGVKELAKQVKS LPVVNYNLLKYICRFLDE
SEG

```

COILS .....
lrp- HHHHTTTTTTGGGHHHHH---TTTTCGGGHHHHHHHHHCHHHHHHHHHHHHHHHHH

SEQ VQSYSGVNMKMSVQNLATVFGPNILRPKVEDPLTIMEGTVVVQQLMSVMISKHDCLPKDA
SEG .....
COILS .....
lrp- HHHHHHHHCCCHHHHHHHHGGGCC.....

SEQ ELQSKPDQGVSNNEIQKATMGLLQNKENNTKDSFSRQCSWDKSESFQRSSMNNGSPT
SEG .....
COILS .....
lrp- .....

SEQ ALSGSKTNSPKNSVHKLDVSRSPPLMVKKNPAPFNKSGIVTNGSFSSSSNAEGLEKTQTP
SEG .....
COILS .....
lrp- .....

SEQ NGSQARRSSSLKVSQKMGTHSVQNGTVRMGILNSDTLGNPTNVRNMSWLPNGYVTLRD
SEG .....
COILS .....
lrp- .....

SEQ NKQKEQAGELGQHNLSTYDNVHQFMMNLDDKQSIDSATWSTSSCEISLPENSNSCRS
SEG .....xxxxxxx
COILS .....
lrp- .....

SEQ STTTCPEQDFFGNFEDPVLGPPQDDLSPRDYESKSDHRSVGGRSSRATSSSDNSETF
SEG xxxxx.....xxxxxxxxxxxxxxxxxxxxx...
COILS .....
lrp- .....

SEQ VGNSSSNHSLVSSLKQEMTKQIEYESRIKSLEQRNLTLETMMSLHDELDQERKK
SEG ..xxxxxxxxxxxxxxxxxxxxx.....
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
lrp- .....

SEQ FTMIEIKMRNAERAKEDAERNDMLQKEMEQQFSTFGELTVEPRRTERGNTIWIQ
SEG .....
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....
lrp- .....

```

Prosites for DKFZphfbr2_62b11.1

PS00001	271->275	ASN_GLYCOSYLATION	PDOC00001
PS00001	342->346	ASN_GLYCOSYLATION	PDOC00001
PS00001	361->365	ASN_GLYCOSYLATION	PDOC00001
PS00001	386->390	ASN_GLYCOSYLATION	PDOC00001
PS00001	407->411	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	547->551	ASN_GLYCOSYLATION	PDOC00001
PS00001	580->584	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	367->371	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	599->603	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	47->50	PKC_PHOSPHO_SITE	PDOC00005
PS00005	309->312	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	388->391	PKC_PHOSPHO_SITE	PDOC00005
PS00005	417->420	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	527->530	PKC_PHOSPHO_SITE	PDOC00005
PS00005	557->560	PKC_PHOSPHO_SITE	PDOC00005
PS00005	646->649	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	213->217	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	348->352	CK2_PHOSPHO_SITE	PDOC00006
PS00006	417->421	CK2_PHOSPHO_SITE	PDOC00006
PS00006	437->441	CK2_PHOSPHO_SITE	PDOC00006
PS00006	465->469	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	516->520	CK2_PHOSPHO_SITE	PDOC00006
PS00006	532->536	CK2_PHOSPHO_SITE	PDOC00006

PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	602->606	CK2_PHOSPHO_SITE	PDOC00006
PS00006	635->639	CK2_PHOSPHO_SITE	PDOC00006
PS00007	43->51	TYR_PHOSPHO_SITE	PDOC00007
PS00007	176->185	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	249->255	MYRISTYL	PDOC00008
PS00008	263->269	MYRISTYL	PDOC00008
PS00008	297->303	MYRISTYL	PDOC00008
PS00008	304->310	MYRISTYL	PDOC00008
PS00008	338->344	MYRISTYL	PDOC00008
PS00008	343->349	MYRISTYL	PDOC00008
PS00008	352->358	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	376->382	MYRISTYL	PDOC00008
PS00008	392->398	MYRISTYL	PDOC00008
PS00008	400->406	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	542->548	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phfbr2_62b11.1)

DKFZphfbr2_62f10

group: intracellular transport and trafficking

DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with strong similarity to mammalian zinc transporter proteins.

The novel proteins is a membrane protein, which should be involved in the transport of Zinc across the cell membrane.

The Zn-T-transporters are membrane proteins that facilitates sequestration of zinc in endosomal vesicles. In the brain, ZnT-3 mRNA seems to be involved in the accumulation of zinc in synaptic vesicles. Zinc (Zn) is an essential element in normal development and metabolism. Recent studies show that in Alzheimer's disease, Zn functions as a double-edged sword, affording protection against Alzheimer's amyloid beta peptide (the major component of senile plaques) at low concentrations and enhancing toxicity at high concentrations by accelerated aggregation of the amyloid beta peptide.

The new protein can find application in modulation of Zinc transport in neuronal cells, thus providing means for a modulation of Alzheimer's amyloid beta peptide plaque formation.

strong similarity to zinc transporter proteins ;
membrane regions: 5

Summary DKFZphfbr2_62f10 encodes a novel 320 amino acid protein with similarity to zinc transporter protein.

The new protein can find clinical application in modulating Zn²⁺ uptake.

strong similarity to zinc transporter proteins

complete cDNA, complete cds, few EST hits

Sequenced by LMU

Locus: unknown

Insert length: 5422 bp

Poly A stretch at pos. 5397, polyadenylation signal at pos. 5381

```
1  GTCTAACTTT  GGAAATATCA  CCCTCATGCT  GTCTTCCCAG  GATGTCTCTC
51  TCCTTAAGTA  AGGGATGTTA  CTTCTGGGAG  GGAATGCAGT  GTTGGGAATC
101 TGAAGACCCA  GCTTTGAGCT  GAATTTGCTT  TGTGATACCT  GGAGAGAAGA
151 CGTGTTTTCT  TGACAACAGC  ACAGTACCTA  GTGAGTTCAA  CAACAACGAC
201 AACAAAGAGC  GCAGCTCATC  CTGGCCGTCA  TGGAGTTTCT  TGAAGAGCGC
251 TATCTTGTTG  ATGATAAAGC  TGCCAAGATG  TATGCTTTCA  CACTAGAAAG
301 AAGGAGCTGC  AAATGAACAC  TTCATAGCAA  TGTGGAACCT  CAACAGAAAC
351 CGGTGAATAA  AGATCAGTGT  CCCAGAGAGA  GACCAGAGGA  GCTGGAGTCA
401 GGAGGGCATGT  ACCACTGCCA  CAGTGGCTCC  AAGCCCACAG  AAAAGGGGGC
451 GAATGAGTAC  GCCTATGCCA  AGTGGAAACT  CTGTTCTGCT  TCAGCAATAT
501 GCTTCATTTT  CATGATTGCA  GAGGTCGTGG  GTGGGCACAT  TGCTGGGAGT
551 CTTGCTGTGT  TCACAGATGC  TGCCACCTC  TTAATTGACC  TGACCACTTT
601 CCTGCTCAGT  CTCTTCTCCC  TGTGGTTGTC  ATCGAAGCCT  CCCTCTAAGC
651 GGTGATGATT  TGGATGGCAC  CGAGCAGAGA  TCCTTGGTGC  CCTGCTCTCC
701 ATCCTGTGCA  TCTGGGTGGT  GACTGGCGTG  CTAGTGTACC  TGGCATGTGA
751 GCGCCTGCTG  TATCCTGATT  ACCAGATCCA  GGCGACTGTG  ATGATCATCG
801 TTTCCAGCTG  CGCAGTGGCG  GCCAACATTG  TACTAACTGT  GGTTTTGAC
851 CAGAGATGCC  TTGGCCACAA  TCACAAGGAA  GTACAAGCCA  ATGCCAGCGT
901 CAGAGCTGCT  TTTGTGCATG  CCCCTGGAGA  TCTATTTTCA  AGTATCAGTG
951 TGCTAATTAG  TGCATTATT  ATCTACTTTA  AGCCAGAGTA  TAAATAGCC
1001 GACCCAATCT  GCACATTCAT  CTTTCCATC  CTGGTCTTGG  CCAGCACCAT
1051 CACTATCTTA  AAGGACTTCT  CCATCTTACT  CATGGAAGGT  GTGCCAAAGA
1101 GCCTGAATTA  CAGTGGTGTG  AAAGAGCTTA  TTTTAGCAGT  CGACGGGGTG
1151 CTGTCTGTGC  ACTGCCTGCA  CATCTGGTCT  CTAACAATGA  ATCAAGTAAT
1201 TCTCTCAGCT  CATGTTGCTA  CAGCAGCCAG  CCGGGACAGC  CAAGTGGTTC
1251 GGAGAGAAAT  TGCTAAAGCC  CTTAGCAAAA  GCTTTACGAT  GCACCTCACT
1301 ACCATTGAGA  TGGAATCTCC  AGTTGACCAG  GACCCGACT  GCCTTTCTG
1351 TGAAGACCCC  TGTGACTAGC  TCAGTCACAC  CGTCAGTTTC  CCAAAATTGA
1401 CAGGCCACCT  TCAAACATGC  TGCTATGCAA  TTTCTGCATC  ATAGAAAATA
1451 AGGAACCAAA  GGAAGAAATT  CATGTCATGG  TGCAATGCAT  ATTTTATCTA
1501 TTTATTAGT  TCCATTCAAC  ATGAAGGAAG  AGGCACTGAG  ATCCATCAAT
1551 CAATTGGATT  ATATACTGAT  CAGTAGCTGT  GTTCAATTGC  AGGAATGTGT
1601 ATATAGATTA  TTCCTGAGTG  GAGCCGAAGT  AACAGCTGTT  TGTAATATC
1651 GGCAATACCA  AATTCATCTC  CCTTCCAATA  ATGCATCTTG  AGAACACATA
1701 GGTAAATTTG  AACTCAGGAA  AGTCTTACTA  GAAATCAGTG  GAAGGACAAA
1751 ATAGTCACAA  AATTTTACCA  AAACATTAGA  AACAAAAAAT  AAGGAGAGCC
1801 AAGTCAGGAA  TAAAAGTGAC  TCTGTATGCT  AACGCCACAT  TAGAATTG
```

```

1851 TTCTCTCACC AAGCTGTAAT GTGATTTTTT TTTCTACTCT GAATTGGAAA
1901 TATGTATGAA TATACAGAGA AGTGCTTACA ACTAATTTTT ATTTACTTGT
1951 CACATTTTGG CAATAAATCC CTCTTATTTT TAAATTTCTA CTTGTTTATT
2001 TCAAAACTTT ATATAATCAC TGTTCAAAAG GAAATATTTT CACCTACCAG
2051 AGTGCTTTAA CACTGGCACC AGCCAAAGAA TGTGGTTGTA GAGACCCAGA
2101 AGTCTTCAAG AACAGCCGAC AAAAACATTC GAGTTGACCC CACCAAGTTG
2151 TTGCCACAGA TAATTTAGAT ATTTACCTGC AAGAAGGAAT AAAGCAGATG
2201 CAACCAATTC ATTCAGTCCA CGAGCATGAT GTGAGCACTG CTTTGTGCTA
2251 GACATTGGGC TTAGCACTGA AACTATAAAG AGGAATCAGA CGCAGCAAGT
2301 GCTTCTGTGT TCTGGTAGCA ACTCAACACT ATCTGTGGAG AGTAACTGA
2351 AGATGTGCAG GCCAACATTC TGGAAA1CCT ATGTCAGTGG GTTTGGTTTG
2401 GAACCTGGAC TTCTGCATTT TTAAAAGTTA CCCAGAGATG CTTCTAAAGA
2451 TGAGCCATAG TCTAGAAGAT TGTCAACCA AGGAGTTCAT TGAGTGGGAC
2501 AGCTAGACAC ATACATTGGC AGTTACAATA GTATCATGAA TTGCAATGAT
2551 GTAGTGGGGT ATAAAAGGAA AGCGATGGAT ATTGCCGGAT GGGCATGGCC
2601 AGTGATGTTT CACGTCATTG AGGTGACAGC TCTGTGGAC TTTGAATTAC
2651 ATATGGAGGC TCTCCAGGAA GACGAAGAAG AGAAGGACAT TCTAGGCAAA
2701 AAGAAAGACTA GGCACAAGGC ACACCTATGT TTGTCGTGTA GCTTTTAGTT
2751 GAAAAAGCAA AATACATGAT GCAAAGAAAC CTCTCCACGC TGTGATTTTT
2801 AAAACTACAT ACTTTTTGCA ACTTTATGGT TATGAGTATT GTAGAGAACA
2851 GGAGATAGGT CTTAGATGAT TTTTATGTTG TTGTCAGACT CTAGCAAGGT
2901 ACTAGAAACC TAGCAGGCAT TAATAATTGT TGAGGCAATG ACTCTGAGGC
2951 TATATCTGGG CCTTGTCAAT ATTTATCATT TATATTTGTA TTTTTTCTG
3001 AAATTTGAGG GCCAAGAAAA CATTGACTTT GACTGAGGAG GTCACATCTG
3051 TGCCATCTCT GCAAAATCAAT CAGCACCAC GAAATAACTA CTTAGCATTC
3101 TGCTGAGCTT TCCCTGCTCA GTAGAGACAA ATATACTCAT CCCCCACCTC
3151 AGTGAGCTTG TTTAGGCAAC CAGGATTAGA GCTGCTCAGG TTCCCAACGT
3201 CTCTGCCAC ATCGGGTCTT CAAAATGGAA AGAATGGTTT ATGCCAAATC
3251 ACTTTTCCCTG TCTGAAGGAC CACTGAATGG TTTTGTTTTT CCATATTTTG
3301 CATAGAGACG CCTAAAGACT AGGTGACTTG GCAAAACACAC AAGTGTTAGT
3351 ATAATTCTTT GCTTCTGCTT CTTTTTGAAA ATCATGTTTA GATTGATTT
3401 TAAGTCAGAA ATTCACGTAA GTGCAGGTAA TCATTATGGA GGGAGATTTG
3451 TGTGTCAAAC AAAGTAATTG TCCCATGGCC CCAGGGTATT TCTGTTGTTT
3501 CCTGAAATTT CTGCTTTTTT AGTCAGCTAG ATTGAAAAC CTGAACAGTA
3551 GATGTTTATA TGGCAAAATG CAAGACAATC TATAAGGGAG ATTTTAAGGA
3601 TTTTGAGATG AAAAAACAGA TGCTACTCAG GGGCTTTATG GACCATCCAT
3651 CAATTCAGAA GTTCTGACTC TCCCATTACC CTTTCCCTGG TGTGGTCAGA
3701 ACTCCAGGTC ACTGGAAGTT AGTGAATCA TGTAGTTGAA TTCTTTACTT
3751 CAAGACATTG TATTCTCTCC AGCTATCAAA ACATTAAATGA TCTTTTATGT
3801 CTTTTTTTTG TTATTGTTAT ACTTTAAGTT CTGGGGTACA TGTGCGGAAC
3851 ATGTAGGTTT GTTACATAGG TATACATGTG CCATGGTGGT TTGCTGCACT
3901 CATCAACCTG TCATCTACAT TCTTTTATGT CTGCTTTTCA AAGCAACACT
3951 CTGTTCTTCT GAGTAGTGAA ATCAGGTCAA CTTTACCACC AGCCTCCATT
4001 TTTAATATGC TTCACCATCA TCCAGCACCT ACTTAAGATT TATCTAGGGC
4051 TCTGTGGTGA TGTAGGACC CATAAAGAA ATTTATGCCT TCCATATGTT
4101 TGGTTACAGA TGGGAAATGG GAATGTTGAA GGACATGAAA GAAAGGATGT
4151 TTACACATTA AGCATCAGTT CTGAAGCTAG ATTGTCTGAG TTTGAATCTT
4201 AGCTCTTCCC TTTATTAGCT CTGTGACCTC GAGCTAGTTA CTTAAATGCT
4251 CTGATCCTCT ATTTCCCTGAT CAGTGAACCC TCCCTATTCA AATGTGTGAG
4301 AGTTTAAATA ATTAGGACAC TTAATAATGT TGGAGCAGTG CATAGCATGT
4351 AGTGTTCACT ACATGTTAAA TGTGTTTTT TATTATGTAC AAACATGTGT
4401 GGGCACAGAA TTTTAAATCA TCTCACTTTT TGAGAAATTT TGAGTTATCA
4451 ACACCGTTCC CACAAGACAG TGGCAAAATT ATTGGTGAGA ATTAACACGC
4501 TGTTTCTCAG AGGAAGCAAT GGAGGCTTGC TGGGATAAAG GCATTTACTG
4551 AGAGGCTGTT ACCTAGTGAG AGTGATGAAT TAATTAAAT AGTCGAATCC
4601 CTTTCTGACT GTCTCTGAAA GCTTCCGCTT TTATCTTTGA AGAGCAGAAT
4651 TGTACGCCCA AGGACATTTA TTAATAAAAA GAACAACCTG CCAGTGCAAT
4701 GAAGGCCAAG TCATAGGTCT CCAAGTCTT ACCCATTTCC TGTGAAATAT
4751 CAAGTCTCTG GCTTTTCTCT GTCACTGAGC CTCAACTTTC TCCGACCGGG
4801 TGCATTCTTT TCTCTGGTTT CTAATTTGCC AGTGGCAAAAT TTGGATCACT
4851 TACTTAATAT CTGTTAAATT TTGTAGCCCA ACAAAGTCTT TTAGCACTGT
4901 GGTGTCAAAA AGAAAAACAC CTCCCAGGCA TATACATTTT ATAGATTCTT
4951 GGAGAATGTT GCTCTCCAGC TCCATCCCA CCAATGAAA TATGATCCAG
5001 AGAGTCTTGC AAAGAGACAA GCCTCATTTT CCACAATTAG CTCTAAAGTG
5051 CCTCCAGGAA ATGATTTTCT CAGCTCATCT CTCTGTATTC CCTGTTTTGG
5101 ATCAGAGGGC AATCTGTTTA AATGACTAAT TACAGAAATC ATTAAGGCA
5151 CCAAGCAAAAT GTCATCTCTG AATACACACA TCCCAAGCTT TACAAATCCT
5201 GCCTGGCTTG ACAGTGATGA GGCCACTTAA CAGTCCAGCG CAGGCGGATG
5251 TTAATAAAAA TAAAAAGGTG ACCATCTGCG GTTTAGTTTT TTAACTTTCT
5301 GATTTACAC TTAACGTCTG TCATTCTGTT ACTGGGCACC TGTTTAAATT
5351 CTATTTTAAA ATGTTAATGA GTGTTGTTTA AAATAAAATC AGGAAAGAGA
5401 GAAAAA AAAA AC

```

BLAST Results

No BLAST result

Medline entries

97121493:
ZnT-3, a putative transporter of zinc into synaptic vesicles.

96203098:
ZnT-2, a mammalian protein that confers resistance to zinc by
facilitating vesicular
sequestration.

Peptide information for frame 2

ORF from 407 bp to 1366 bp; peptide length: 320
Category: strong similarity to known protein

```

1 MYHCHSGSKP TEKGANEYAY AKWKLCSASA ICFIFMIAEV VGGHIAGSLA
51 VVTDAAHLLI DLTSFLLSLF SLWLSSKPPS KRLTFGWHRRA EILGALLSIL
101 CIWVVVTGVLV YLACERLLYP DYQIQATVMI IVSSCAVAAN IVLTVVLHQR
151 CLGHNHKEVQ ANASVRAAFV HAPGDLFQSI SVLISALIIY FKPEYKIADP
201 ICTFIFSLIV LASTITILKD FSILLMEGVP KSLNYSVGVE LILAVDGVLS
251 VHCLHIWSLT MNQVILSAHV ATAASRDSQV VRREIAKALS KSFTMHSLTI
301 QMESPVQDQP DCLFCEDPCD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_62f10, frame 2

PIR:S70632 zinc transporter ZnT-2 - rat, N = 1, Score = 884, P = 1.5e-88

TREMBL:MMU76007_1 gene: "ZnT-3"; product: "ZnT-3"; Mus musculus zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 772, P = 1.1e-76

TREMBL:HSU76010_1 gene: "ZnT-3"; product: "ZnT-3"; Human putative zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds., N = 1, Score = 742, P = 1.6e-73

TREMBL:MMUZNT02_1 gene: "ZnT-3"; product: "zinc transporter"; Mus musculus zinc transporter (ZnT-3) gene, complete cds., N = 1, Score = 715, P = 1.2e-70

TREMBL:CET18D3_3 gene: "T18D3.3"; Caenorhabditis elegans cosmid T18D3, N = 1, Score = 699, P = 5.9e-69

>PIR:S70632 zinc transporter ZnT-2 - rat
Length = 359

HSPs:

Score = 884 (132.6 bits), Expect = 1.5e-88, P = 1.5e-88
Identities = 171/326 (52%), Positives = 230/326 (70%)

```

Query:      2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61
             ++CH+      +E  A+ KL  ASAIC +FMI E++GG++A SLA++TDAHLL D
Sbjct:     34 HYCHAQKDSGSHPNSEKQARRKLYVASAICLVFMIGEIIIGGYLAQSLAIMTDAHLLTD 93

Query:     62 LTSFLLSLFSLWLSSKPPSKRLTFGWHRRAEILGALLSILCIWVVVTGVLVYLACERLLYPD 121
             S L+SLFSLW+SS+P +K + FGW RAEILGALLS+L IWVVVTGVLVYLA +RL+ D
Sbjct:     94 FASMLISLFLWVSSRPATKTMNFGWQRAEILGALLSVLSIWVVVTGVLVYLAVQRLISGD 153

Query:    122 YQIQATVMIIVSSCAVAANIVI.TVVLHQRCLGHNH-----KEVQANASVRAAFVHAPG 174
             Y+I+  M+I S CAVA NI++ + LHQ  GH+H      + Q N SVRAAF+H  G
Sbjct:    154 YEIKGDTMLITSGCAVAVNIIMGLALHQSGHGHSHGHSHEDSSQQQONPSVRAAFIHVVG 213

Query:    175 DLFQSIISVLISALIIYFKPEYKIADPICTFIFSLIVLASTITILKDFSILLMEGVKPSLN 234
             DL QS+ VL++A IYFKPEYK DPICTF+FSILVL +T+TIL+D ++LMEG PK ++
Sbjct:    214 DLLQSVGVLVAAAYIYFKPEYKYVDPICTFLFSILVLTTLTILRDVILVLMEGTPKGVGD 273

Query:    235 YSGVKELILAVDGVLSVHCLHIWSLTMNQVILSAHVATAASRDSQVVRREIAKALSKSFT 294
             ++ VK L+L+VDGV ++H LHIW+LT+ Q +LS H+A A + D+Q V +      L  F
Sbjct:    274 FTTVKNLLLSVDGVEALHSLHIWALTVAQPVLSVHIAIAQNVDQAQVAVLKVARDRLQGKFN 333

```

```
Query: 295 MHSLTIQMESPVDPDPCLFCEDPCD 320
      H++TIQ+ES + C C+ P +
Sbjct: 334 FHTMTIQIESYSEDMKSCQECQPSE 359
```

Pedant information for DKF2phfbr2_62f10, frame 2

Report for DKF2phfbr2_62f10.2

```

[LENGTH] 320
[MW] 35053.51
[pI] 6.48
[HOMOL] PIR:S70632 zinc transporter Znt-2 - rat 3e-84
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YMR243c] 2e-16
[FUNCAT] 13.01 homeostasis of metal ions [S. cerevisiae, YMR243c] 2e-16
[FUNCAT] 08.19 cellular import [S. cerevisiae, YMR243c] 2e-16
[FUNCAT] 11.07 detoxification [S. cerevisiae, YMR243c] 2e-16
[FUNCAT] 07.04.01 metal ion transporters (cu, fe, etc.) [S. cerevisiae, YMR243c]
2e-16
[FUNCAT] 08.04 mitochondrial transport [S. cerevisiae, YOR316c] 3e-13
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YOR316c] 3e-13
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YDR205w] 4e-07
[PIRKW] transmembrane protein 2e-30
[PIRKW] mitochondrial inner membrane 6e-12
[PIRKW] mitochondrion 6e-12
[PIRKW] membrane protein 1e-11
[SUPFAM] zinc transporter Znt-2 2e-30
[SUPFAM] membrane protein czcD 1e-11
[PROSITE] MYRISTYL 4
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 1
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 4
[PROSITE] ASN_GLYCOSYLATION 2
[KW] TRANSMEMBRANE 5
[KW] LOW_COMPLEXITY 8.12 %

```

```
SEQ      MYHCHSGSKPTEKGANEYAYAKWKLCSAIAICFIFMIAEVVGGHIAGSLAVVTDAAHLLI
SEG      .....                               .xxx
PRD      cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhh
MEM      ..mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm
```

```
SEQ      DLTSFLLSLPSLWLSSKPPSKRLTFGWHRRAEILGALLSILCIWVTVGVLYVYLACERLLYP
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
MEM      MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....
```

```
SEQ      DYQIQATVMIIVSSCAVAANI VLTVV LHQRC LGHNHKEV QANASVR AAFV HAPGDL FQS I
SEG
PRD      cccccccceee hhhhhh hhhhhh hhhhhccc cccccccc chhhhh hhhhhh hhchhh
MEM      mmmmmmmmmmmmmmmmmmmmmmmmm . . . . . mmmmmmmmmmmmmmmmmmmmm
```

```
SEQ      SVLSALISLIIFYKPEYKIADPICTFIFSLVLASTITILKDPSILLMEGVPKSLNYSVGVE
SEG      .
PRD      .
MEM      hhhhhhhhhhccccceccchhhhhhhhhhhhhhhhhhhhhheeecccccccchhhhh
          .MMMMMMMMMMMMMMMMMMM.
```

```
SEQ      LILAVDGVLSVHCLHIWLSLTMNQVILSAHVATAARDSQVVRREIAKALSKSFTMHSLTI
SEG
PRD      hhhhhhceeeccceeeecchhhheeeeccccchhhhhhhhhhhhhhhccccee
MEM
```

```
SEQ      QMESPV DQDP DCLFCEDPCD
SEG      .....
PRD      eecccccccccccccccccc
MEM      .....
```

Prosites for DKFZphfbr2 62f10.2

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00001	234->238	ASN_GLYCOSYLATION	PDOC00001
PS00004	81->85	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005

WO 01/12659

PCT/IB00/01496

PS00005	80->83	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00007	13->21	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	228->234	MYRISTYL	PDOC00008
PS00013	125->136	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_62f10.2)

DKFZphfbr2_62n10

group: brain derived

DKFZphfbr2_62n10 encodes a novel 541 amino acid protein with similarity to *Plasmodium vivax* reticulocyte-binding protein 1.

The novel protein contains one Leucine Zipper, involved in protein-protein-interaction. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to reticulocyte-binding protein

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="13"

Insert length: 3522 bp

Poly A stretch at pos. 3503, polyadenylation signal at pos. 3479

```
1 GGGGCGTGTT GCGGGGATTC TGAACGCTGC CATGGCTCAG ACCGTGTAGA
51 ATGTTACATT GTCGCTCACT CTGCCCATCA CGTGCCACAT TTGCTTGGGG
101 AAGGTACGTC AGCCTGTTCAT ATGCATCAAC AACCATGTAT TTTGTTTCGAT
151 TTGTATTGAT TTGTGGTTGA AGAATAATAG CCAGTGTCCA GCTTGCAGAG
201 TCCCCATCAC TCCTGAAAAT CCTTGCAAAG AAATTATAGG AGGAACAAGT
251 GAAAGTGAAC CTATGCTAAG CCATACGGTC AGGAAGCATC TTCGGAAAAAC
301 TAGACTTGAA TTAATACACA AAGAATATGA GGACGAAATA GATTGTTTAC
351 AGAAAGAAGT AGAAGAGCTT AAGAGTAAAA ATCTCAGCTT GGAGTCACAG
401 ATCAAAGCTA TTCTGGATCC TTAAACCTTG GTGCAGGGCA ACCAAAATGA
451 AGACAAACAT CTAGTCACAG ATAATCCAAG TATAATTAAC CCAGAAACTG
501 TAGCAGAGTG GAAGAAAAAA CTCAGAACAG CTAATGAAAT CTATGAAAAA
551 GTGAAAGATG ATGTGGATAA GCTAAAGGAG GCAATAAAAA AATTGAAATT
601 GGAATAATGGT GGTCTGGTGA GGGAGAATTT ACGACTGAAG GCTGAAAGTTG
651 ATAACAGATC ACCTCAAAAG TTTGGAAAGT TTGCAGTTGC TGCTCTTCAG
701 TCCAAAGTAG AACAGTATGA GCGTGAAACC AATCGCCTCA AGAAAGCCCT
751 GGAACGAAGT GATAAGTATA TAGAGGAACT AGAATCTCAA GTTGCACAGC
801 TAAAAAATTC AAGTGAAGAG AAAGAGGCTA TGAATTCCAT TTGCCAGACA
851 GCACCTTCTC CAGATGGCAA AGGGAGCAAA GGCACTGAGG AGGATGTGGT
901 GTCAAGAAGT CAAGGCGATA GTGCCAGAAA GCAGCCTGGC TCATCCACCT
951 CCACTTCTTC TCACCTAGCG AAGCCTTCCA GCAGCAGACT GTGTGACACC
1001 AGTTCTGCAG GGCAGGAAAG TACCAGCAAA GCAGACCTTA ACTGTTCTAA
1051 GAACAAGAGC CTATATCAAG AACAGGTAGA AGTAATGTTA GATGTGACAG
1101 ATACAAGTAT GGATACTTAT TTGGAAAGAG AATGGGGGAA TAAACCAAGT
1151 GACTGTGTAC CCTACAAAGA TGAAGAACTT TATGATTTTC CAGCTCCTTG
1201 TACTCCTTTG TCCCTTAGTT GCCTTCAGCT CAGTACTCCA GAAATAGAG
1251 AGAGCTCTGT GGTCCAAGCA GGAGGTTCCT AAAAGCACTC AAACCATCTC
1301 AGAAAAATGG TGTGTGATGA TTTTGTGAT TCTTCAAATG TTTCTAATAA
1351 AGATTCTTCA GAAGATGATA TAAGTAGAAG TGAAATGAG AAGAAATCAG
1401 AATGTTTTC TTCCACAAAG ACAGGATTTT GGGACTGTTG TTCCACAAGC
1451 TATGCCCAAA ACTTAGATTT TGAAAGTTCA GAGGGGAACA CGATAGCAAA
1501 TTCTGTTGGA GAAATATCTT CAAAATTGAG TGAGAAATCA GGCTTATGTT
1551 TATCCAAAAG GTTGAATCTT ATTGCTCTT TTGAAATGAA CCGGACAAGA
1601 ACATCCAGTG AAGCATCGAT GGATGCTGCT TACCTTGACA AAATCTCTGA
1651 GTTGGATTCA ATGATGTCAG AGTCAGACAA CAGCAAGAGC CCTTGTAATA
1701 ACGGTTTTAA GTCACGGAT TTGGATGGGT TATCAAAGTC ATCTCAAGGC
1751 AGTGAATTTT TTGAGGAACC TGATAAGTTG GAAGAAAAAA CTGAGCTAAA
1801 CCTTTCCAAA GGTTCCTCTA CTAATGATCA GTTAGAAAAA GGAAGTGAAT
1851 GGAACCCAC TTCTTTTTTT TCTCCTCTCT CCATCTGACC AAGAAATGAA
1901 TGAAGATTTT TCACTCCATT CCACTTCTTG TCCAGTAACT AATGAAATCA
1951 AACCCCAAG CTGCTTGTTT CAGACAGAGT TTTCCAGGG CATTTTGTTA
2001 AGCAGTTCAC ATCGACTATT GGAAGATCAA AGATTGGGT CATCTTGTG
2051 TAAGATGTC TCAGAGATGC ACAGTCTTCA TAACCACCTT CAGTCTCCTT
2101 GGTCTACTTC CTTTGTGCCT GAAAAGAGGA ATAAAAATGT GAATCAATCA
2151 ACAAAGAA AAATCCAGAG CAGCCTTTCC AGTGCCAGCC CATCAAAAGC
2201 AACTAAAAAT TGACTCATT GAAAGGTGTC ATTTGTGGTT TTGCTCTGAG
2251 AGAAATAGAA AAGTTGTTAA AGTTACCTTT TTTCTCATA AAAGTTCTAT
2301 ACAAATGGA ATTGATAATC TTAGTCAAG TATCAAGTCA GGATGGTGA
2351 TTAACCTGTA CCCAGAAATC TTATTGTTCA TTTTGAAAG ACTTTGTTCT
2401 TTTCATTTTT ATTTGGGAGT CTTTGTGACC AGAGAAGTTA GGGAGGAGGT
2451 TATTTTGTG TTTTGGGGTT GGTGGTTGG TTGGTTTGT TTTTGGTTT
2501 GTTTTTTAC TGAATTTGAT ATGTATCTCG GTTGATATA CATTGTTTT
2551 TTAATAAATG TTATTTAACT GTTAGATACA GTGGCCTGTT GATAAGCCCC
2601 ACTTGTCTTC AGAAGTTGGA TTTCTTAAAT AAAACTTTTA GTGTGTCTA
```

```

2651 TACACTGCTC AATAAGACAC TTGAGTTTAA GCTTTTCCCA GGGTGGAAAT
2701 TATTTTACCT GTCCCTTTTT ATTTATGTTT AGTGATGGCC TAGTTTTTCT
2751 GCAGGGCCAT GATGGAGAAA TAGCACTCTA GCCTTAGTCC AATATTGATT
2801 TACTTTCTTT TTTTAGGTTT TATGTATATG TTTGCATTTT TTAGCATTGT
2851 GTTTTGTCCA GTTTTGTGAA AATGTTCTGC TAGTATGAAA GAAACATTT
2901 TCTATATGAA GACATTTGTT TTAATGTTAGG TAGCTTACAT TTTCTCTCT
2951 GCGTGTGTGT GTATGTGTGT AAAATCAGAA ATTTAGCATA CTATGGAAAG
3001 AAGGCATGGA GCACCTGGGT TTAGAGGAAC CTAACACATC ATAGCTTCAT
3051 TGTTCCAGAT GTAACAGGTT TGAAGAGCT CATCGCCAAG TTCTTGATCC
3101 ACTTGCAATC CAGGGGAGTT CTCTTTGAG TAGTATGTTT CTGTTTGCA
3151 TGTTCTCTGT CTTTGTGGA ACTATGCATG GTAGCATTTT TGCTTGCTGT
3201 GTTTTCCATA CTTAAGAAAA AGAGGTTTCA GTTGGCTGAT AGAATATCTT
3251 TTATGTAGGA CAAACTTTT CTGTGAAGAG TGTGAGGGG GTGAAGATAG
3301 GTAAGAGGTA AGCACAAATT TTAATTAGG CTCTGAAAAA GTGTATTGTT
3351 CTAAACGTAT TTGGTATGCC TATATAGGTC TTTAAAAATG GGTGTGTATG
3401 CTGTTTAATG TGCACTGAAC ATTTTACATT AATATTGTAC TGTTTTACAT
3451 TAATACTGCA TGCTTTTCTA TGTGAATTGA ATAAAGAATG TCATAAGCAC
3501 TGGAAAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS658254 from database EMBL:
human STS SHGC-11774.
Score = 1643, P = 8.0e-67, identities = 345/355

Entry HS513217 from database EMBL:
human STS SHGC-14656.
Score = 1193, P = 5.8e-46, identities = 241/244

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 263 bp to 1885 bp; peptide length: 541
Category: similarity to known protein

```

1  MLSHTVRKHL RKTRLELLHK EYEDEIDCLQ KEVEELKSKN LSLESQIKAI
51  LDPLTLVQGN QNEDKHLVTD NPSIINPETV AEWKKKLRTA NEIYEKVKDD
101 VDKLKEANKK LKLENGGLVR ENLRKAEVD NRSPOKFGFR AVAALQSKVE
151 QYERETNRLK KALERSDKYI EELESQVQL KNSSEEKAM NSICQTALSA
201 DGKSGSGSEE DVVSKNQGDS ARKQPSSTS SSSHAKPSS SRLCDTSSAR
251 QESTSKADLN CSKNKDLYQE QVEVMDVTD TSMPTYLERE WGNKPSDCVP
301 YKDEELYDFP APCTPLSLSC LQLSTPENRE SSVVQAGGSK KHSNHLRKL
351 FDDFCDSNV SNKDSSEDDI SRSENEKKSE CFSSTKTGFW DCCSTSYAQN
401 LDFESSEGMT IANSVGEISS KLSEKSGLCL SKRLNSIRSF EMNRTRTSSE
451 ASMDAAYLDK ISELDMMSE SDNSKSPCNN GFKSLDLGL SKSSQGSSEFL
501 EEPDKLEKT ELNLSKGLT NDQLENGSEW KPTSFSPPLS I

```

BLASTP hits

Entry A42771 from database PIR:
reticulocyte-binding protein 1 - Plasmodium vivax
Score = 127, P = 3.7e-08, identities = 68/300, positives = 145/300

Entry RBPl_PLAVB from database SWISSPROT:
RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
Score = 127, P = 3.9e-08, identities = 68/300, positives = 145/300

Entry MMDSPPG_1 from database TREMBL:
gene: "DSPP"; product: "dentin sialophosphoprotein"; Mus musculus DSPP
gene
Score = 160, P = 5.2e-08, identities = 87/373, positives = 146/373

Alert BLASTP hits for DKFZphfbr2_62n10, frame 2

No Alert BLASTP hits found

Report for DKFZphfbr2_62n10.2

```

SEQ      MLSHTVRKHKRKTRLELLHKEYEDEIDCLQKEVEELKSNKLSLESQIKAILDLPLTLVQGN
SEG
PRD      cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhcccccccccc
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      QNEDKHLVTDNPSIINPETVAEWKKKLRATANIEYIEYKVDVVKLEANKKLLKLENGGLVR
SEG
PRD      cccceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceee
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      ENLRLKAEVDNRSPOKFGFRAVALQSKVEQYERETNRLKKALERSDKYIEELESQVAQL
SEG
PRD      ehhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      KNSSEEKAMNSICQTALSADGKSGKSGEEDVVSKNQGDSARKQPGSSTSSSHLAKPSS
SEG
PRD      hcchhhhhhhhhhhhhhhccccccccccccceeeccccccccccccccccccccccccccccccccc
COILS    CCCCCC.....

SEQ      SRLCOTSSARQUESTSKADLNCNKNDLYQEVEVMDLVTDTSMDTYLEREWGNKPSDCVP
SEG
PRD      x.....
COILS    cccccccccccccccccccccccccchhhhhhhhhccccccccccchhhhhhhhhcccccccccc

SEQ      YKDEELYDFPAPCTPLSLCQLSTPENRESSVUQAGGSKKSHNLRKLVDFDQCSSNV
SEG
PRD      cccccccccccccccccccccceeeccccccccceeecccccccccccccccccccccccccccccc
COILS    .....

SEQ      SNKDSSEDDIRSNEKKSECFSSSTKTGFWDCCSTSYAQNLFESSEGNTIANSVGEISS
SEG
PRD      cccccccchhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

SEQ      KLSEKSGCLCKSLRNSIRSFEMNRTTSSSEASMDAAYLDKISELDSMMSSESDNSKSPCNN
SEG
PRD      cccccccchhhhhccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccc
COILS    .....

SEQ      GFKSLDLGLSKSSQGFLEFPDKLEEKTELNLSKGSGLTNQLENGSEWKPTSFPSPLS
SEG
PRD      .xxxxxxxxxxxxxxxxx.
COILS    cccccccccccccccccceccccchhhhhhhhhccccccccccccccccccccccccccccccccc

SEQ      I
SEG      .
PRD      c
COILS    c

```

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	182->186	ASN_GLYCOSYLATION	PDOC00001
PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001

PS00001	359->363	ASN_GLYCOSYLATION	PDOC00001
PS00001	443->447	ASN_GLYCOSYLATION	PDOC00001
PS00001	513->517	ASN_GLYCOSYLATION	PDOC00001
PS00001	526->530	ASN_GLYCOSYLATION	PDOC00001
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	156->159	PKC_PHOSPHO_SITE	PDOC00005
PS00005	166->169	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	240->243	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	361->364	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	419->422	PKC_PHOSPHO_SITE	PDOC00005
PS00005	423->426	PKC_PHOSPHO_SITE	PDOC00005
PS00005	431->434	PKC_PHOSPHO_SITE	PDOC00005
PS00005	436->439	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	183->187	CK2_PHOSPHO_SITE	PDOC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	324->328	CK2_PHOSPHO_SITE	PDOC00006
PS00006	361->365	CK2_PHOSPHO_SITE	PDOC00006
PS00006	365->369	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00006	373->377	CK2_PHOSPHO_SITE	PDOC00006
PS00006	414->418	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	462->466	CK2_PHOSPHO_SITE	PDOC00006
PS00006	469->473	CK2_PHOSPHO_SITE	PDOC00006
PS00007	294->302	TYR_PHOSPHO_SITE	PDOC00007
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	226->232	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	408->414	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	489->495	MYRISTYL	PDOC00008
PS00008	517->523	MYRISTYL	PDOC00008
PS00013	310->321	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_62n10.2)

DKFZphfbr2_62o17

group: metabolism

DKFZphfbr2_62o17.2 encodes a novel 282 amino acid protein with weak similarity to the apolipoprotein E receptor.

The new protein contains a leucine zipper for protein-protein interaction, and three LDL-receptor class A domain (LDLRA_1) patterns. In LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands.

The new protein can find application in modulation of cholesterol binding and transport by LDL-receptors and LDL-binding proteins

similarity to apolipoprotein E receptor

complete cDNA, complete cds, start at Bp 56 matches kozak consensus
ANCatg EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1260 bp

Poly A stretch at pos. 1240, polyadenylation signal at pos. 1218

```

1 GGGGGATAAG AGAGCGGTCT GGACAGCGCG TGGCCGGCGC CGCTGTGGGG
51 ACAGCATGAG CGGCGGTTGG ATGGCGCAGG TTGGAGCGTG GCGAACAGGG
101 GCTCTTGGGCC TGGCGTGCT GCTGCTGCTC GGCCTCGGAC TAGGCCTGGA
151 GGCCGCCGCG AGCCCGCTTT CCACCCCGAC CTCGTGCCAG GCCGCAGGCC
201 CCAGCTCAGG CTCGTGCCCA CCCACCAAGT TCCAGTGCCG CACCAGTGGC
251 TTATGCTGCG CCCTCACCTG GCGCTGCGAC AGGGACTTGG ACTGCAGCGA
301 TGGCAGCGAT GAGGAGGAGT GCAGGATTGA GCCATGTACC CAGAAAGGGC
351 AATGCCACCC GCCCCCTGGC CTCCCTGCC CCTGCACCGG CGTCAGTGAC
401 TGCTCTGGGG GAACTGACAA GAACTGCGC AACTGCAGCC GCCTGGCCTG
451 CCTAGCAGGC GAGCTCCGTT GCACGCTGAG CGATGACTGC ATTCCACTCA
501 CGTGGCGCTG CGACGGCCAC CCAGACTGTC CCGACTCCAG CGACGAGCTC
551 GGGCTGTGAA CCAATGAGAT CCTCCCGGAA GGGGATGCCA CAACCATGGG
601 GCCCCTGTG ACCCTGGAGA GCGTCACCTC TCTCAGGAAT GCCACAACCA
651 TGGGGCCCCC TGTGACCCTG GAGAGTGTC CCTCTGTCGG GAATGCCACA
701 TCCTCCTCTG CCGGAGACCA GTCTGGAAGC CCAACTGCCT ATGGGGTTAT
751 TGCAGCTGCT GCGGTGCTCA GTCAAGCCT GGTACCGGCC ACCCTCCTCC
801 TTTTGTCTTG GCTCCGAGCC CAGGAGCGCC TCCGCCCACT GGGGTTACTG
851 GTGGCCATGA AGGAGTCCCT GCTGTGTGCA GAACAGAAGA CCTCGCTGCC
901 CTGAGGACAA GCACTTGCCA CCACCGTCAC TCAGCCCTGG GCGTAGCCGG
951 ACAGGAGGAG AGCAGTGATG CGGATGGGTA CCCGGGCACA CCAGCCCTCA
1001 GAGACCTGAG CTCTTCTGGC CACGTGGAAC CTCGAACCCG AGCTCCTGCA
1051 GAAGTGGCCC TGGAGATTGA GGGTCCCTGG AACTCCCTA TGGAGATCCG
1101 GGGAGCTAGG ATGGGGAACC TGCCACAGCC AGAACCAGG GGCTGGCCCC
1151 AGGCAGCTCC CAGGGGGTAG GACGGCCCTG TGCTTAAGAC ACTCCTGCTG
1201 CCCCCTCTGA GGGTGGCGAT TAAAGTTGCT TCACATCCTC AAAAAAAAAA
1251 AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 56 bp to 901 bp; peptide length: 282
Category: similarity to known protein
Classification: unset
Prosite motifs: LDLRA_1 (67-90)
LDLRA_1 (67-90)
LDLRA_1 (145-168)

LEUCINE_ZIPPER (17-39)

```

1 MSGGWMQVQ AWRTGALGLA LLLLLGLGLG LEAAASPLST PTSQAAGPS
51 SGSCPPTKFQ CRTSGLCVPL TWRCRDRLDC SDGSDEEECR IEPCTQKGQC
101 PPPPGLPCPC TGVSDCSGGT DKLRNCSRL ACLAGELRCT LSDDCIPLTW
151 RCDGHPDCPD SSDELGCGTN EILPEGDATT MGPPVTLESV TSLRNATTMG
201 PPVTLESVPS VGNATSSSAG DQSGSPATYG VIAAAVLSA SLVTATLLLL
251 SWLRAQERLR PLGLLVAMKE SLLSEQKTS LP

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_62o17, frame 2

TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene., N = 1, Score = 733, P = 1.5e-72

PIR:JE0237 apolipoprotein E receptor 2 precursor - mouse, N = 2, Score = 290, P = 1.1e-26

TREMBL:HSZ75190.1 product: "apolipoprotein E receptor 2 906"; H.sapiens mRNA for apolipoprotein E receptor 2, N = 1, Score = 279, P = 1.8e-23

>TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial cds; and unknown gene.
Length = 260

HSPs:

Score = 733 (110.0 bits), Expect = 1.5e-72, P = 1.5e-72
Identities = 157/276 (56%), Positives = 178/276 (64%)

```

Query:      6 MAQVGAWRTGALGLALLLLGLGLGLEAAASPLSTPTSQAAGPSSGSCPPTKFQCRTSG 65
            MA+ GA R ALGL L LL GL GLEAA +P T Q +G + SCP FQC TSG
Sbjct:      1 MARGGAGRAVALGLVLRLLFLRTGLEAAPAHT--RVQVSGSRADSCPTDTFQCLTSG 58

Query:     66 LCVPLTWRCRDRLDCSDGSDEEECRIEPCTQKGCPPPPGLPCPCTGVSDCSGGTDKKLR 125
            CVPL+WRCD D DCSGDSDEE+CRIE C Q GQC P LPC C +S CS +DK L
Sbjct:     59 YCVPLSWRCDGQDCSDGSDEEDCRIECAQNGQCQPQSALPCSCDNISGCSDVSDKNL- 117

Query:    126 NCSRLACLAGELRCTLSDDCIPLTWRCDGHPDCPDSSDELGCGTNEILPEGDATTMGPPV 185
            NCSR C EL C L D CIP TWRCDGHPDC DSSDEL C T+
Sbjct:    118 NCSRPPCQESLHCILDDVCIPHWTWRCDGHPDCLDSSDELSCDTD-----T 163

Query:    186 TLESVTSLRNATTMGPPVTLESVPSVGNATSSSAGDQSGSPATYGVIAAAVLSASLVTA 245
            ++ + NATT T+E+ S N T +SAGD S +P+AYGVIAAA VLSA LV+A
Sbjct:    164 EIDKIFQEENATTTRISTMENETSFRNVFTSAGDSSRNPSAYGVIAAGVLSAILVSA 223

Query:    246 TLLLLSWLRAQERLRPLGLLVAMKESLLSEQKTS 281
            TLL+L LR Q L P GLLVA+KESLLSE+KTS
Sbjct:    224 TLLILLLRLRGQGYLPPPGLLVAVKESLLSERKTS 259

```

Pedant information for DKF2phfbr2_62o17, frame 2

Report for DKF2phfbr2_62o17.2

```

[LENGTH]      282
[MW]           28991.19
[pI]           4.61
[HOMOL]        TREMBL:AF110520.6 product: "NG29"; Mus musculus major histocompatibility
                complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein,
                BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes,
                complete cds; Sacm21 gene, partial cds; and unknown gene. 5e-55
[BLOCKS]       BL01209 LDL-receptor class A (LDLRA) domain proteins
[SCOP]         d1ajj_ 7.1.1.1.1 Ligand-binding domain of low-density lipoprotei 2e-10

```

```

SEQ      MSGGWMNAQVQJAWRTGALGLALLLLGLGLGLEAAASPLSTPTPSAQAGPSSGSCPPTKFQ
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccce
MEM      .....

SEQ      CRTSGLCVP17LTWRCDRLDCSDGSDDEECRIEPTCKQKQCPPPGLPCPTGVSDCSGGT
SEG      .....XXXXXXXXXXXX
PRD      ecccccccccecccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      DKKL17RNC17SRLACLAGELR17CTLSDDCI17PLTWRC17DGH17PDC17PDSSDELGC17GT17NI17LP17EGDATT
SEG      .....
PRD      ccccccccccccccccccecccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      MGPPVTLESV17TSLRNATMGPPVTLESVPSVG17NATSSAGDQSGSP17AYGVIAAAVLSA
SEG      .....XXXXXXXXXX
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh
MEM      .....MMMMMM

SEQ      SLVTATLLLSWLRAQERL17PLGLIVAMKES17LLSEQKTS17LP
SEG      xxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhcccccc
MEM      mmmmmmmmmmm

```

```

PS01209      67->90    LDLRA_1                PDOC00929
PS01209      67->90    LDLRA_1                PDOC00929
PS01209      145->168  LDLRA_1                PDOC00929
PS00029      17->39    LEUCINE ZIPPER         PDOC00029

```

HMM_NAME	TNFR/NGFR cysteine-rich region	
HMM	*CpeGtYtD.WNHvpqClpC.trCePEMGQYmvqPCTwTQNT.VC*	
	CP+ ++ + + + C+P RC+ + + C+ ++ + + C	
Query	54	CPPTKFFQRTS--GLCVPLTWRCDR--DL----DCSDGSDDEEEC 89


```

HMM_NAME      Low-density lipoprotein receptor domain class A
HMM            *tTCeGPDEFQCgSGeMRCIFMsWvCDGdpDCeDWSDEWPeNChp*
               C P +FQC+++ C+P+ W+CD D DC D+SDE E+C+
Query         52  GSCP-PTKFCRTSG-LCVPLTWRCDRDLDCSDGSDE--EECRI    91

54.99 (bits) f: 130 t: 169 Target: dkfzphfbr2_62ol7.2 similarity to apolipoprotein E
receptor
Alignment to HMM consensus:
Query         *tTCeGPDEFQCgSGeMRCIFMsWvCDGdpDCeDWSDEWPeNChp*
               C + E +C + CIP+ W+CDG PDC D SDE ++C+
dkfzphfbr2    130  LACL-AGELRCTLSD-DCIPLTWRCDGHPDCPDSSDE--LGCCT    169

```

DKFZphfbr2_64a15

group: nucleic acid management

DKFZphfbr2_64a15 encodes a novel 255 amino acid protein with strong similarity to inorganic pyrophosphatases

Inorganic pyrophosphatase (EC 3.6.1.1) (PPase) is the enzyme responsible for the hydrolysis of pyrophosphate (PPi) which is formed as the product of the many biosynthetic reactions that utilize ATP. All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity.

The new protein can find application as a new enzyme for biotechnologic processes.

strong similarity to inorganic pyrophosphatases

unspliced Intron 212-256 see EST HS1190948

Sequenced by Qiagen

Locus: unknown

Insert length: 1188 bp

Poly A stretch at pos. 1170, polyadenylation signal at pos. 1151

```

1 GGGGGTTGGG GACCAGTGCA GGGACCGGGT CGGCGCGTGC TATGGCCCTG
51 TACCACACTG AGGAGCGCGG CCAGCCCTGC TCGCAGAATT ACCGCTCTT
101 CTTAAGAAT GTAACGTGTC ACTACATTTC CCCCTTTCAT GATATTCCTC
151 TGAAGGTGAA CTCTAAAGAG GACACTGAGG CTCAAGGCAT TTTTATAGAC
201 TTGCTAAGA TCTGGAAAT GGCATTCTTA TGAAGAAAGC ACGAAATGAT
251 GAATATGAGA ATCTGTTTAA TATGATTGTA GAAATACCTC GGTGGACAAA
301 GGCTAAAATG GAGATTGCCA CCAAGGAGCC AATGAATCCC ATTAACAAT
351 ATGTAAAGGA TGGAAAGCTA CGCTATGTGG CGAATATCTT CCCTTACAAG
401 GGTATATAT GGAATTATGG TACCTCCCTC CAGACTTGGG AAGATCCCCA
451 TGA AAAAGAT AAGAGCACGA ACTGCTTTGG AGATAATGAT CCTATTGATG
501 TTTGCGAAAT AGGCTCAAAG ATTCTTTCTT GTGGAGAAGT TATTCATGTG
551 AAGATCCTTG GAATTTTGGC TCTTATTGAT GAAGGTGAAA CAGATTGGAA
601 ATTAATTGCT ATCAATGCGA ATGATCCTGA AGCCTCAAAG TTTCATGATA
651 TTGATGATGT TAAGAAGTTC AAACCGGGTT ACCTGGAAGC TACTCTTAAT
701 TGGTTTAGAT TATGTAAGGT ACCAGATGGA AAACCAGAAA ACCAGTTTGC
751 TTTTAAATGA GAATTCAAAA ACAAGGCTTT TGCTCTTGAA GTTATTAAT
801 CCACATCATCA ATGTTGGAAA GCATTGCTTA TGAAGAACTG TAATGGAGGA
851 GCTACAAATT GCACAAACGT GCAGATATCT GATAGCCCTT TCCGTGTCAC
901 TCAAGAGGAA GCAAGATCAT TAGTTGAATC GGTATCATCT TCACCAATA
951 AAGAAAGTAA TGAAGAAGAG CAAGTGTTGC ACTTCCTTGG CAAGTGATTG
1001 AAACATCTGA AATTCTGCTG TCAAGATTCC CATCTCTAAG GACTCCAAGA
1051 CTCTTTTTC CCAAGTGCTA GAGACAAGGG GGTCTATGAG CATTCTACTGA
1101 CTCTCTGTTA AAACCTCATT TTTCAAAC TTTTGAGCTA TGCAATATAT
1151 AATAAACAG TAAGAAATTT AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

Entry HSPPASEMR from database EMBL:
H.sapiens partial mRNA for pyrophosphatase.
Score = 1706, P = 1.6e-70, identities = 342/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 230 bp to 994 bp; peptide length: 255
Category: strong similarity to known protein
Classification: unset
Prosites motifs: PPASE (85-92)

```

1 MKKARNDEYE NLFNMIVEIP RWTAKMEIA TKEPMNPIKQ YVKDGKLRVY
51 ANIFPYKGYI WNYGTLPTW EDPHEKDKST NCFGDNDPID VCEIGSKILS
101 CGEVIHVKIL GILALIDEGE TDWKLIAINA NDPEASKFHD IDDVKKFKPG
151 YLEATLNWFR LCKVPDGKPE NQFAFNGEFK NKAFALVLIK STHQCWKALL
201 MKNCNGGATN CTNVQISDSP FRCTQEEARS LVESVSSSPN KESNEEQVW
251 HFLGK

```

BLASTP hits

Entry IPYR_KLULA from database SWISSPROT:
 INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-
 HYDROLASE) (PPASE).
 Score = 689, P = 6.0e-68, identities = 128/248, positives = 170/248

Entry A45153 from database PIR:
 inorganic pyrophosphatase (EC 3.6.1.1) - bovine
 Score = 862, P = 2.8e-86, identities = 146/226, positives = 190/226

Entry AF085600.1 from database TREMBLNEW:
 gene: "Nurf-38"; product: "inorganic pyrophosphatase NURF-38";
 Drosophila melanogaster inorganic pyrophosphatase NURF-38 (Nurf-38)
 gene, complete cds.
 Score = 731, P = 2.1e-72, identities = 134/248, positives = 177/248

Entry PWB1 from database PIR:
 inorganic pyrophosphatase (EC 3.6.1.1) - yeast (Saccharomyces
 cerevisiae)
 Score = 688, P = 7.7e-68, identities = 133/251, positives = 174/251

Alert BLASTP hits for DKFZphfbr2_64a15, frame 2

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
 (PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 731, P =
 2.4e-72

>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
 PHOSPHO- HYDROLASE) (PPASE).
 Length = 290

HSPs:

Score = 731 (109.7 bits), Expect = 2.4e-72, P = 2.4e-72
 Identities = 134/248 (54%), Positives = 177/248 (71%)

```

Query:      7 DEYENLNFNMIVEIPRWTAKMEIATKEPMNPIKQYVKDGKLRVYANIFPYKGYIWNNGTL 66
            +E + ++NM+VE+PRWT AKMEI+ K PMNPIKQ +K GKLR+VAN FP+KGYIWNNG L
Sbjct:     40 NEEKTIYNMVVEVPRWTNAKMEISLKTPMNPIKQDIKKGKLRVANCFFPHKGYIWNNGAL 99

Query:     67 PQTWEDPHEKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWKLI 126
            PQTWE+P + ST C GDNDPIDV EIG ++ G+V+ VK+LG ALIDEGETDWK+I
Sbjct:    100 PQTWENPDHIEPSTGCKGDNDPIDVIEIGYRVAKRGDVLKVKVLGQFALIDEGETDWKII 159

Query:     127 AINANDPEASKFHDIDDVKKFKPGYLEATLNWFRCKVPDGKPEHQFNGEFKNKAFAL 186
            AI+ NDP ASK +DI DV ++ PG L AT+ WF++ K+PDGKPEHQFNG+KN FA
Sbjct:    160 AIDVNDPLASKVNDIADVQYFPGLLRATVEWFKIYKIPDGKPEHQFNGDAKNADFAN 219

Query:     187 EVIKSTHQCWKALLMKNCNGGATNCTNVQISDSPFRCTQEEARS-LVESVSSSPNKESNE 245
            +I TH+ W+ L+ ++ G+ + TN+ +S +EEA L E+ +E ++
Sbjct:    220 TIIAETHKFWQNLVHQSASGSISTTNITNRNSEHVIPKEEAELAEAPDGGQVEEVSD 279

Query:     246 EEQVWHFL 253
            WHF+
Sbjct:    280 TVDTWHFI 287

```

Peptide information for frame 3

ORF from 42 bp to 230 bp; peptide length: 63
 Category: strong similarity to known protein
 Classification: unset

```

1 MALYHTEERG QPCSQNYRLF FKNVTGHIYS PFHDIPLKVN SKEDTEAQGI
51 FIDLSKIWM AFL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64a15, frame 3

SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1)
(PYROPHOSPHATE PHOSPHO- HYDROLASE) (PPASE)., N = 1, Score = 118, P = 8.8e-07

PIR:A45153 inorganic pyrophosphatase (EC 3.6.1.1) - bovine, N = 1,
Score = 113, P = 3.1e-06

TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase";
Homo sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds., N
= 1, Score = 106, P = 1.8e-05

>SWISSPROT:IPYR_DROME INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE
PHOSPHO- HYDROLASE) (PPASE).
Length = 290

HSPs:

Score = 118 (17.7 bits), Expect = 8.8e-07, P = 8.8e-07
Identities = 23/43 (53%), Positives = 29/43 (67%)

Query: 1 MALYHTEERGQPCSONYRLFFKNVTGHIYISPFHDIPLKVNKE 43
MALY T E+G S +Y L+FKN G+ ISP HDIPL N ++
Sbjct: 1 MALYETVEKGAKNSPSYSLYFKNKGCVISPMHDIPLYANEK 43

Pedant information for DKFZphfbr2_64a15, frame 2

Report for DKFZphfbr2_64a15.2

[LENGTH] 255
[MW] 29177.34
[pI] 5.67
[HOMOL] TREMBLNEW:AF108211_1 product: "cytosolic inorganic pyrophosphatase"; Homo
sapiens cytosolic inorganic pyrophosphatase mRNA, partial cds. 2e-93
[FUNCAT] 01.04.01 phosphate utilization [S. cerevisiae, YBR011c] 9e-73
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR011c] 9e-73
[FUNCAT] 02.99 other energy generation activities [S. cerevisiae, YMR267w] 1e-58
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR267w] 1e-58
[FUNCAT] 1 genome replication, transcription, recombination and repair [M.
genitalium, MG351] 1e-06
[FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0124] 2e-06
[BLOCKS] BL00387D
[BLOCKS] BL00387C
[BLOCKS] BL00387B
[BLOCKS] BL00387A
[SCOP] dlwgja_2.29.5.1.1 Inorganic pyrophosphatase [baker's yeas 1e-113
[EC] 3.6.1.1 Inorganic pyrophosphatase 7e-92
[PIRKW] mitochondrion 3e-57
[PIRKW] hydrolase 7e-92
[PIRKW] homodimer 2e-71
[SUPFAM] inorganic pyrophosphatase 7e-92
[PROSITE] PPASE 1
[KW] Alpha_Beta
[KW] 3D
[KW] LOW_COMPLEXITY 6.27 %

SEQ MKKARNDHEYENLFNMIVEIPRWTKAKMEIATKEPMNPIKQYVKGDKLRYVANIFPYKGYI
SEG
lhukBEGGGCEEEEEETTTbCBCEETTTTTTCEEECEETTECBCCBTTbTbT

SEQ WNYGTLPTQWEDPHEKDKSTNCFGDNDDPIDVCEIGSKILSCGEVIHVKILGILALIDEGE
SEG
lhukB CEEETTTTbTbTTEETTTTECCCBCEEECCCECCCTTTEEEEEEEEEETTTbT

SEQ TDWKLIAINANDPEASKFHDIDDVKKFKPGYLEATLNWFLRCKVPDGKPNQFAFNGEFK
SEG
lhukB CEEEEEEEEETTTTGGGCCCHHHHHHTTTTHHHHHHHHHHHC GGCCCCC BC GGCCB

SEQ NKAFALEVIKSTHCWKALLMKNCNGGATNCTNVQISDSPFRCTQEEARSLVESVSSPN
SEGxxxxxxx
lhukB CHHHHHHHHHHHHHHHHHCTTTTTTCCCBTTTTTTT

PCT/IB00/01496

(No Pfam data available for DKF2phfbr2 64a15.3)

DKFZphfbr2_64c16

group: brain derived

DKFZphfbr2_64a16.2 encodes a novel 101 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: /map="745_A_2; 756_F_2; 842_C_2"

Insert length: 1866 bp

Poly A stretch at pos. 1848, polyadenylation signal at pos. 1829

```
1 GGGCGCGGCG CCGGAGGAGG AAGTGGTGAG GTTGTGTGCTC CTTCAGCGCC
51 TATCGCTGGC TCTTGGGGCG CAGAGAGGGG CCGCAGTCTC CGCGGCTGCG
101 TCGAGCTCCC TTGCAGTCCC CTCCATGTTC CCCGGCGCCA CTACTCCCCT
151 TCCTAAGGCC GCCGCTTACC CCGGGGTCTA TGGAAAGTAAT GGAAGGACCC
201 CTCACCTGGG CTCATCAACA GAGCAGACGA GCAGACCGTT TATTAGCTGC
251 AGGCAAATAC GAAGAGGCTA TTTCTTGTC CAAAAGGCT GCAGCATATC
301 TTTCTGAAGC CATGAAGCTG ACACAGTCAG AGCAGGCTCA TCTTCACTG
351 GAATTGCAAA GGGATAGCCA TATGAAACAG CTCCTCCTCA TCCAAGAGAG
401 ATGGAAAAGG GCCCAGCGTG AAGAAAGATT GAAAGCCCAG CAGAACACAG
451 ACAAGGATGC AGCTGCCCAT CTTCAGACAT CTCACAAACC CTCTGCAGAG
501 GATGCAGAGG GCCAGAGTCC CCTTCTCAG AAGTACAGCC CTTCCACAGA
551 GAAATGCCCTG CCTGAGATTC AGGGGATCTT TGACAGGGAT CCAGACACAC
601 TACTTTATTT ACTTCAGCAA AAGAGTGAGC CAGCAGAGCC ATGTATTGGA
651 AGCAAAGCCC CAAAAGATGA TAAACAATT ATAGAGGAGC AGGCAACCAA
701 AATTGCAGAT TTGAAGAGGC ATGTGGAAAT CCTTGTGGCT GAGAATGAAA
751 GATTAAGGAA AGAAAATAAA CAACTAAAGG CTGAAAAGGC CAGACTTCTA
801 AAAGGTCCAA TAGAAAAGGA GCTGGATGTA GATGCTGATT TTGTAGAAAC
851 GTCAGAGTTA TGGAGCTTGC CACCACATGC AGAACTGCT ACAGCCTCCT
901 CAACCTGGCA GAAGTTCGCA GCAAACTACTG GGAAGGCCAA GGACATTCCA
951 ATCCCCAATC TTCCTCCCTT GGATTTTCCA TCTCCAGAAC TTCCTTTAT
1001 GGAGCTCTCT GAGGATATTC TGAAGGACT TATGAATAAT TAAATGGAA
1051 GGGCACAGAA AAGGGGAAAA GAGGAAATAA TACAGTAATC GTTAATCCAG
1101 CAAAAGAAAA TGAAAAGGGA AAACACATA GAAGGGTAAT CCCGGAATG
1151 CTTCTCTGGG TGGACTGTGG GAGCAGAGGC ATTGCCAGGA CTTGGGAAAC
1201 AGTCACTGTG AARTGCGCTG CGTATCTCAT TCACTCACTT CAGCTAATGA
1251 CTCCGACTTG GCAGACGCTA AACTCATGGA GGTTCGGTTT CTCCTGATAC
1301 AAACCAAAATG GCTACCTGGA AGAATTTCTT TCAAGCAACA GTTATTTTTC
1351 TTATCTTCAG GGTAAAATG TATAAAAGTT ATGTGTAATT AATCTATAAT
1401 GCCATAAATG ATAATGCAAA ACCTAAATAA TATGGTGGCC GGAGGGGCTG
1451 CCTTATATTT GAAACATGCT TTCTATCATG CATTGACTGT ATGCATTTTG
1501 TTAATGCACA TTCTGTTTGT TTAAGGTGTG TGAGATACAC ACCTTTCTAG
1551 ATGAACTAT ATGTGCCACA CTTTGCACTA CTCATAATGA TAACCTCAAG
1601 ACTATCAGAA GAAATATTTA AATTCCATT TTATGAAGAA AGGAACCAAA
1651 TTATTTATGCT TTTTAAACA AATTACCAGT TTACATAATT AATCAGGGTG
1701 CATTTTAACT TCTAACTTCG TTTATTGTAT AATGCATCAT TTGAAAATAC
1751 CAAGGAGGAA ATACCCTTTG TTTTAAATGA TGCAAGAGTG GACGTAATGC
1801 TAGTTGGCAG TATTTTATTG TAAGAAATCA ATAAAGTAAT TGTGTTTTAA
1851 AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS286143 from database EMBL:

human STS WI-6844.

Score = 1460, P = 3.4e-61, identities = 292/292

Medline entries

No Medline entry

1 GAAPEEEVVR LLLLQRLSLA LGAQRGA AVS AAASSSLAVP SMFPGATTPL
51 PKAAAYPGVY GSNGRTPQPG SSTEQTSRPF ISCRQIRRGY FLSQKGCSIS
101 F

No BLASTP hits available

No Alert BLASTP hits found

```

1  MEVMEGPLNL AHQQSRRADR LLAAGKYEEA ISCHKKAAAY LSEAMKLTQS
51  EQAHLSELQL RDSHMKQLLL IQERWKRAER EERLKAAQNT DKDAHAALQT
101 SHKPSADEAS GQSPLOQSYK PSTEKCLPEI QGIFDRDPDT LLLYLLQKSE
151 PAEPCIGSKA PKDDKTIIEE QATKIADLKR HVEFLVAENE RLRKENKQLK
201 AEKARLLKFG IEGELDVADP FVTELSWLSI PPHAETATAS STWQKFAANT
251 GKARDIPINL LPPLDFPSFE PLMELSEDLI LKGLMNN

```

No BLASTP hits available

No Alert BLASTP hits found

.....

```
[LENGTH]      101
[MW]           10469.94
[pI]           10.18
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      29.70 %
```

(No Pfam data available for DKFZphfbr2 64c16.2)

Report for DKFZphfbr2_64c16.3

```

[LENGTH]      287
[MW]           32343.79
[pI]           5.61
[PROSITE]      LEUCINE_ZIPPER 2
[KW]           All_Alpha
[KW]           COILED_COIL      14.98 %

SEQ  MEVMEGPLNLAHQSSRRADRLAAGKYEEAISCHKKAAAYLSEAMKLTQSEQAHLSELEQ
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ  RDSHMKQLLLIQERWKRAQREERLKAQQNTDKDAAHLQTSKPSAEDAEGQSPLSQKYS
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ  PSTEKCLPEIQGIFDRDPDTLLYLLQKSEPAEPCIGSKAPKDDKTIIEEQATKIADLKR
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCCC

SEQ  HVEFLVAENERLRKENKQLKAEKARLLKGPKEKELDVDADFVETSELWSLPPHAETATAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ  STWQKFAANTGKAKDIPINLPPLDFPSPELPLMELSEDILKGLMNN
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

Prosites for DKFZphfbr2_64c16.3

```

PS00029  178->200  LEUCINE_ZIPPER      PDOC00029
PS00029  185->207  LEUCINE_ZIPPER      PDOC00029

```

(No Pfam data available for DKFZphfbr2_64c16.3)

DKFZphfbr2_64c4

group: brain derived

DKFZphfbr2_64c4 encodes a novel 467 amino acid protein with similarity to A. thaliana T08I13.5

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A. thaliana T08I13.5

complete cDNA, complete cds, EST hits
on genomic level encoded by AC005043 11 exons

Sequenced by Qiagen

Locus: unknown

Insert length: 1559 bp

Poly A stretch at pos. 1540, no polyadenylation signal found

```
1 TGGGACCGCC GGAAGTTTCT GCCCGGCTT TGCAGGACG GGGGAGTGGT
51 AGTGGGGGCT GCAGCTGCCG GACCCAGGCG CGATGGCTAC GGGCGCGGAT
101 GTACGGGACA TTCTAGAACT CGGCGGTCCA GAAGGGGATG CAGCCTCTGG
151 GACCATCAGC AAGAAGGACA TTATCAACCC GGACAAGAAA AAATCCAAGA
201 AGTCCTCTGA GACACTGACT TTCAAGAGGC CCGAGGGCAT GCACCGGGAA
251 GTCTATGCTT TGCTCTACTC TGACAAGAAG GATGCACCCC CACTGTCTACC
301 CAGTGACACT GGCCAGGGAT ACCGTACAGT GAAGGCCAAG TTGGGTCCA
351 AGAAGGTGCG GCCTTGGGAG TGGATGCCAT TCACCAACCC GGCCTGCAAG
401 GACCGAGCAA TGTTCTTCCA CTGGCGACGT GCAGCGGAGG AGGGCAAGGA
451 CTACCCCTTT GCCAGTTTCA ATAAGACTGT GCAGGAGCCT GTGTACTCGG
501 AGCAGGAGTA CCAGCTTTAT CTCCACGATA ATGCTTGGAC TAAGGCAGAA
551 ACTGACCACC TCCTTGACCT CAGCCGCCGC TTTGACCTGC GTTTTGTGT
601 TATCCATGAC CGGTATGACC ACCAGCAGTT CAAGAAGCGT TCTGTGGAAG
651 ACCTGAAGGA GCGGTACTAC CACATCTGTG CTAAGCTTGC CAACGTGCGG
701 GCTGTGCCAG GCACAGACCT TAAGATACCA GTATTGTATG CTGGGCACGA
751 ACGACGGCGG AAGGAACAGC TTGAGCGTCT CTACAACCGG ACCCCAGAGC
801 AGGTGGCAGA GGAGGAGTAC CTGCTACAGG AGCTGCGCAA GATTGAGGCC
851 CGGAAGAAGG AGCGGGAGAA ACGCAGCCAG GACCTGCAGA AGCTGATCAC
901 AGCGGCAGAC ACCACTGCAG AGCAGCGGCG CACGGAACGC AAGGCCCCCA
951 AAAAGAAGCT ACCCCAGAAA AAGGAGGCTG AGAAGCGGCG TGTTCTGAG
1001 ACTGCAGGCA TCAAGTTTCC AGACTTCAAG TCTGCAGGTG TCACGCTGCG
1051 GAGCCAACGG ATGAAGCTGC CAAGCTCTGT GGGACAGAAG AAGATCAAGG
1101 CCCTGGAACA GATGCTGCTG GAGCTTGGTG TGGAGCTGAG CCGCACACT
1151 ACGGAGGAGC TGGTGCACAT GTTCAATGAG CTGCGAAGCG ACCTGGTGCT
1201 GCTCTACGAG CTCAGCAGG CCGTGCCCAA CTGCGAGTAT GAGCTGCAGA
1251 TGCTGCGGCA CCGTCATGAG GCACTGGCCC GGGCTGCTGT GCTAGGGGGC
1301 CCTGCCACAC CAGCATCAGG CCCAGGCCCG GCCTCTGCTG AGCCGGCAGT
1351 GTCTGAACCC GGACTTGGTC CTGACCCCAA GGACACCATC ATTGATGTGG
1401 TGGGCGCACC CCTCACGCC AATTCGAGAA AGCGACGGGA GTCGGCCTCC
1451 AGCTCATCTT CCGTGAAGAA AGCCAAGAA CCGTGAGAGG CCCCACGGGG
1501 TGTGGGCGAC GCTGTTATCT AAATAGAGCT GCTGACTTGG AAAAAAAAAA
1551 AAAAAAAAAA
```

BLAST Results

Entry AC005043 from database EMBL:
Homo sapiens clone NH0576N21; HTGS phase 1, 5 unordered pieces.
Score = 1506, P = 4.6e-244, identities = 316/330

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 83 bp to 1483 bp; peptide length: 467

1	MATGADVDDI	LEGGPEGADA	ASGTISKDDI	INPDKKSSKK	SSETLTFFKRP
5	EGMHREYVAL	LYSDKKOAPP	LPSPDQGGY	KTPVKALGSK	KRPWKVMFKP
10	TNPARKDGMG	FFHRRRAAEE	GKDPYFAFN	KQVQEPVYE	QEYQLYHON
15	AWTKAETHDL	FDSRRRDLER	FVVIHRYDTH	KQCFKRSVED	LKERYIHLHA
20	KLANVARVPE	TDKLTVPVDA	GHERRRKEQL	ERLYNTRPEQ	VAAEYLLOQE
25	LKRIEARKKE	RRKRSQDLK	LITAADTTEE	KRTERKAPK	LKIPKKEAKE
30	KPAVPETTEL	KHFPDKSAGV	TLRSGRMKLP	SSVGGQKLM	LEQMLLELGV
35	ELSPTTTEGI	VMMFDELNSD	LVLYLEKQA	CANCYELYKA	LRHREALAR
40	AGVLGGPAPT	AVSGKGPASAE	PAVSEPLGLP	DPKDTI IDVV	GAPLTPNSRK
45	RRESASSSST	VKKAKKPP			

```

SEQ      MATGADVDRDLELGGPEGDAASGTISKDDIINPDKKKSKSSETLTFKRPEGMHREVYAL
SEG      .XXXXXXXXXXXXXXXXXXXX
PRD      cccceeeeeeeccccccccccccccccccccccccccccccccccccchhhhhhhh

SEQ      LYSDDKDAPLLPSDTGQGYRTVAKLGSKKVRPWKMPFTNPARKDGAMFFHWRRAAE
SEG      .
PRD      hhhccccccccccccccccccccccccccccccccccccccccchhhhhhhhhc

SEQ      GKDYPFARFNKTQEPVYSEQEYQLYLHONAWTKAETHLFDLSRRFDLRFVVIHTRYDH
SEG      .
PRD      cccccccccccccccccchhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhccceeeeeecccc

SEQ      QQFKKRSVDLKERYYHICAKLANVRAPGTDLKIPIVFDAGHERRRKEQLERLYNRTPEQ
SEG      .
PRD      chhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhcchh

SEQ      VAEEYLLQELRKIEARKEREKRSQDLQLLITAADTAEQRRTERKAPKKKLPOKKEAE
SEG      .XXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      KPAPETAGIKFPDFKSAGVTLRSQRMKLPSVSGKKIKALEQMLLELGVLSPTPEEL
SEG      .XXX

```

```
PRD      hccccccccccccccccceehhhhhhhccccccchhhhhhhhhhhhhhhhhccccchh
SEQ      VHMFMELRSLDVLVLYELQKACANCEYELQMLRRHEALARAGVLGGPATPASGPGPASAE
PRD      .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
SEQ      hhhhhhhccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccc
SEG      PAVSEPLGDPDKDTIIDVVGAPLTNPSRKRRRESASSSSSVKKAKKP
SEG      xxxxxxxx.....xxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccceeecccccccccccccccccccccccccccccccc
```

Prosites for DKFZphfbr2_64c4.2

PS000001	130->134	ASN_GLYCOSYLATION	PDOC000001
PS000002	412->416	GLYCOSAMINOGLYCAN	PDOC000002
PS000004	35->39	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	39->43	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	184->188	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	451->455	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	26->29	PKC_PHOSPHO_SITE	PDOC000005
PS000005	38->41	PKC_PHOSPHO_SITE	PDOC000005
PS000005	46->49	PKC_PHOSPHO_SITE	PDOC000005
PS000005	63->66	PKC_PHOSPHO_SITE	PDOC000005
PS000005	82->85	PKC_PHOSPHO_SITE	PDOC000005
PS000005	89->92	PKC_PHOSPHO_SITE	PDOC000005
PS000005	164->167	PKC_PHOSPHO_SITE	PDOC000005
PS000005	284->287	PKC_PHOSPHO_SITE	PDOC000005
PS000005	321->324	PKC_PHOSPHO_SITE	PDOC000005
PS000005	324->327	PKC_PHOSPHO_SITE	PDOC000005
PS000005	448->451	PKC_PHOSPHO_SITE	PDOC000005
PS000005	460->463	PKC_PHOSPHO_SITE	PDOC000005
PS000006	3->7	CK2_PHOSPHO_SITE	PDOC000006
PS000006	26->30	CK2_PHOSPHO_SITE	PDOC000006
PS000006	132->136	CK2_PHOSPHO_SITE	PDOC000006
PS000006	139->143	CK2_PHOSPHO_SITE	PDOC000006
PS000006	153->157	CK2_PHOSPHO_SITE	PDOC000006
PS000006	187->191	CK2_PHOSPHO_SITE	PDOC000006
PS000006	273->277	CK2_PHOSPHO_SITE	PDOC000006
PS000006	277->281	CK2_PHOSPHO_SITE	PDOC000006
PS000006	355->359	CK2_PHOSPHO_SITE	PDOC000006
PS000006	435->439	CK2_PHOSPHO_SITE	PDOC000006
PS000007	131->139	TYR_PHOSPHO_SITE	PDOC000007
PS000007	227->235	TYR_PHOSPHO_SITE	PDOC000007
PS000007	116->125	TYR_PHOSPHO_SITE	PDOC000007
PS000008	14->20	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2_64c4.2)

DKF2phfbr2_64h6

group: brain derived

DKF2phfbr2_64h6 encodes a novel 176 amino acid protein with similarity to predicted yeast proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to S.pombe SPBC337.09 and S.cerevisiae YER044c

complete cDNA, complete cds accoring to YER044c/SPBC337.09,
start at Bp 111, EST hits

Sequenced by Qiagen

Locus: /map="14"

Insert length: 1212 bp

Poly A stretch at pos. 1192, polyadenylation signal at pos. 1168

```
1 GGGCTGGAGC TGTCTGGGG GAGCTTGTTC GCGGCAGCGG CTGCTGCTGC
51 CACTGCTGTG CTGGGGGGCC GGTGCCAGG CAAAAGCCCC TCCCACGTTT
101 GAGGGGAGTC ATGAGCCGTT TCCTGAATGT GTTAAGAAGT TGGCTGGTTA
151 TGGTGTCCAT CATAGCCATG GGAACACGCG TGCAGAGCTT CCGAGACCAC
201 ACTTTTCTCT ATGAAAAGCT CTACACTGGC AAGCCAAACC TTGTGAATGG
251 CCTCCAAGCT CGGACCTTTG GGATCTGGAC GCTGCTCTCA TCAGTGATCC
301 GCTGCTCTG TGCCATTGAC ATTCACAACA AGACGCTCTA TCACATCACA
351 CTCTGGACCT TCCTCCTTGC CTTGGGGCAT TTCCTCTCTG AGTTGTTTGT
401 CTATGGAAGT GCAGCTCCCA CGATTGGCGT CCTGGCAGCC CTGATGGTGG
451 CAAGTTTCTC CATCCTGGGT ATGCTGGTCG GGCTCCGCTA TCTAGAAGTA
501 GAACCAAGTAT CCAGACAGAA GAAGAGAAAC TGAGGCCAGC ATTATCACCT
551 CCAGGACTTT CTCGTTTTCC ACCTTGCCCA TCTTCTTCTT TCGTCGTCTC
601 TCCCCTTTAA TTTCTTTTCT ATTCATCAT CTGCCCTTTT ACTCACCTTT
651 AGCCTCTTTT TTTAATTTTT AAAATTTAAA GATATGCATA CTGAAAAGTA
701 TATAACATGT ACGTACAATT TAAAGAATAA TTTTAAAGTG AATACTACGT
751 AACTCCATCC AAGTCAAGAA ATTGCCAGCT TCTCGGAAGC CCACTGTGTC
801 TCCTTCCCCT ACCTGCAACC TCTTCCAGGC TCCCTTTTCC AGCCTTCCCC
851 TTTTTCCTT TATTTCATC GCCTTGATTT GACTTGTGTG GTGGGAACAT
901 GTGAACATATG AAACCTTAAAC CTGCTGCCCA CCCAGAGCAG CTGTGACCAA
951 GGGCTGCCTC AAGGGTGTGT CCACGCAGGT TGGGCTCCTC TCTGTGCTG
1001 GACCCAAGAC TCTGAACCTT CCAAGGGACA GGCAGTTCTT CTGAGAAGGG
1051 CTCCCCTGTG TGTGAGCAAG ACCACAGCTC TCCTTCTATC TACAGATGCA
1101 TGAGGGTGTG AAGAGTCTGG GCTGTTTTTA GACCTTCTGG TCAGCTGTAT
1151 TTGTGTAAACA ACTTTGTAA TAAATAGAAA AACCTCTGCT TCAAAAAAAA
1201 AAAAAAAAAA AA
```

BLAST Results

Entry G38566 from database EMBL:

SHGC-64295 Human Homo sapiens STS genomic, sequence tagged site.

Score = 1398, P = 1.4e-56, identities = 284/288

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 0 bp to 530 bp; peptide length: 177

Category: similarity to unknown protein

Classification: unclassified

```
1 AGAVLGELVC GSGCCCHCCA GGPVARQKAL PRLRGVMSRF LNVLRSLVM
51 VSIIAMGNLT QSRDHTFLY EKLYTGKPNL VNGLOARTFG IWTLLSSVIR
101 CLCAIDHNK TLYHITLWTF LLALGHFLSE LFVYGTAAPT IGVLAFLMVA
```

151 SFSILGMLVG LRYLEVEPVSRQKKRN

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64h6, frame 3

TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein"; S.pombe chromosome II cosmid c337., N = 1, Score = 224, P = 1.4e-18

PIR:S50547 hypothetical protein YER044c - yeast (Saccharomyces cerevisiae), N = 1, Score = 192, P = 3.4e-15

>TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein"; S.pombe chromosome II cosmid c337.
Length = 136

HSPs:

Score = 224 (33.6 bits), Expect = 1.4e-18, P = 1.4e-18
Identities = 49/113 (43%), Positives = 74/113 (65%)

Query: 42 NVLRSLVMSVSIAMGNTLQSFDRHTFLYEKLYTGKPNLVNGLQARTFGIWTLLSSVIRC 101
+++ W V+VS+ A+ NT+QSF L +++Y+ N VNGLQ RTFGIWTLLS+++R
Sbjct: 11 SLVAKWNVVSVAALENTVQSFLTPK-LTKRVYSNT-NEVNGLQGRTEFGIWTLLSAIVRF 68

Query: 102 LCAIDIHNKTLHYHITLWTFLLALGHFLSELFVYGTAAPTIGVLAPLMVASFSI 154
CA I N +Y + T+ LA HFLSE ++ T G+L+P++V++ SI
Sbjct: 69 YCAYHITNPDVYFLCQCTYYLACFHLSEWLLFRITNLGPGLLSPIVSTVSI 121

Pedant information for DKFZphfbr2_64h6, frame 3

Report for DKFZphfbr2_64h6.3

[LENGTH] 176
[MW] 19359.31
[pI] 9.53
[HOMOL] TREMBL:SPBC337_9 gene: "SPBC337.09"; product: "conserved hypothetical protein";
S.pombe chromosome II cosmid c337. 2e-17
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YER044c] 7e-16
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 7.39 %

SEQ AGAVLGELVCGSGCCCHCCAGGPVARQKALPRLRGVMSRFLNVLRSLVMSVSIAMGNTL
SEGXXXXXXXXXXXXX.....
PRD cccceeeeeeeccccccccccccccccccccchhhhhhhhhhhheeecccccc
MEMMMMMMMMMMMMMMMMM.....

SEQ QSFDRHTFLYEKLYTGKPNLVNGLQARTFGIWTLLSSVIRCLCAIDIHNKTLHYHITLWTF
SEG
PRD cccccchhhhhhhhhccccccccccccccccchhhhhhhhhhhccccceehhhhh
MEM
SEQ LLALGHFLSELFVYGTAAPTIGVLAPLMVASFSILGMLVGLRYLEVEPVSRQKKRN
SEG
PRD hhhhhhhhhhhhhccccccccccccceehhhhhhhhhhhheeecccccccccc
MEMMMMMMMMMMMMMMMMM.....

(No Prosite data available for DKFZphfbr2_64h6.3)

(No Pfam data available for DKFZphfbr2_64h6.3)

DKF2phfbr2_64j18

group: Intracellular transport and trafficking

DKF2phfbr2_64j18.1 encodes a novel 180 amino acid protein nearly identical to the microsomal signal peptidase 23 kd subunit of canis familiaris, gallus gallus and C. elegans.

The new protein is identical to canine and chicken microsomal signal peptidase 23 kd subunit. The canine microsomal signal peptidase is a protein complex comprised of five subunits (25, 22/23, 21, 18, and 12 kDa). The 23kDa subunit is tightly associated with the 18- and 21-kDa subunits, that are integral membrane proteins.

The new protein can find application in modulation of protein transport into microsomal compartments and as a tool for proteomic analysis.

strong similarity to dog signal peptidase (EC 3.4.99.-)

complete cDNA, complete cds, potential start at Bp 109, EST hits,

Sequenced by Qiagen

Locus: unknown

Insert length: 690 bp

Poly A stretch at pos. 666, polyadenylation signal at pos. 646

```

1 GCCGGAACGC GCGCACCGCA GACGGCGCGG ATCGCAGGGA GCCGGTCCGC
51 CGCCGGAACG GGAGCCTGGG TGTGCGTGTG GAGTCCGGAC TCGTGGGAGA
101 CGATCGCGAT GAACACGGTG CTGTCGCGGG CGAACTCACT GTTCGCCTTC
151 TCGCTGAGCG TGATGGCGGC GCTCACCTTC GGCTGCTTCA TCACCACCGC
201 CTTCAAAGAC AGGAGCGTCC CGGTGCGGCT GCACGTCTCG CGGATCATGC
251 TAAAAAATGT AGAAGATTTC ACTGGACCTA GAGAAAGAAG TGATCTGGGA
301 TTTATCACAT CTGATATAAC TGCTGATCTA GAGAATATAT TTGATTGGAA
351 TGTTAAGCAG TTGTTTCTTT ATTTATCAGC AGAATATTCA ACAAATAATA
401 ATGCTCTGAA CCAAGTTGTC CTATGGGACA AGATTGTTTT GAGAGGTGAT
451 AATCCGAAGC TGCTGCTGAA AGATATGAAA ACAAATATT TTTCTTTGA
501 CGATGGAAAT GGTCTCAAGG GAAACAGGAA TGTCACCTTG ACCCTGTCTT
551 GGAACGTCGT ACCAAATGCT GGAATTCTAC CTCTTGTGAC AGGATCAGGA
601 CACGTATCTG TCCCATTTCC AGATACATAT GAAATAACGA AGAGTTATTA
651 AATTATTCTG AATTGAAAC AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

89034208:

cDNA-derived primary structure of the glycoprotein component of canine microsomal signal peptidase complex.

Peptide information for frame 1

ORF from 109 bp to 648 bp; peptide length: 180

Category: strong similarity to known protein

Prosite motifs: TONB_DEPENDENT_REC_1 (1-58)

RGD (148-151)

```

1 MNTVLSRANS LFAFSLSVMA ALTFGCFITT AFKDRSPVVR LHSVRLMKLN
51 VEDFTGPRER SDLGFITSDI TADLENIFDW NVKQLFLYLS AEYSTKNNAL
101 NQVVLWDKIV LRGDNPkLLL KDMKTKYFFF DDGNGLKGNR NVTTLTSLWNV
151 VPNAGILPLV TSGHVSVPF PDTYEITKSY

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64j18, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_64j18, frame 1

Report for DKFZphfbr2_64j18.1

[LENGTH] 180
[MW] 20253.39
[pI] 8.66
[HOMOL] PIR:A31788 signal peptidase (EC 3.4.99.-) (SPC 22/23) - dog le-100
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YLR066w] 6e-15
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YLR066w] 6e-15
[PIRKW] transmembrane protein 2e-92
[PIRKW] glycoprotein 2e-92
[PIRKW] hydrolase 2e-92
[PROSITE] RGD 1
[PROSITE] MYRISTYL 2
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 1
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta
[KW] SIGNAL_PEPTIDE 32

SEQ. MNTVLSRANSLFAFSLSVMAALTFGCFITTAFAKDRSVVRLHVSRIMLKNVEDFTGPRER
PRD cccccchhhhhhhhhhhhhhhhhhhhhheccccceehhhhhhhhhhhhhcccccc

SEQ. SDLGFIITSDITADLENIIDWNVKQLFLYLAEYSTKNNALNQVVLWDKIVLRGDNPKLLL
PRD cccccchhhhhhhccccccchhhhhhhhhhhhhhhccccceeeeeeeccccchhhh

SEQ. KDMKTKYFFFDGNGLGKGNRNVTLTLSWNVVFNAGILPLVTGSGHVSVPFPDTEITKSY
PRD hhccccceeeeeccccccccccccceeeeeccccceeeeeccccceeeeecccccccc

Prosites for DKFZphfbr2_64j18.1

PS00001	141->145	ASN_GLYCOSYLATION	PDOC00001
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	135->141	MYRISTYL	PDOC00008
PS00013	16->27	PROKAR_LIPOPROTEIN	PDOC00013
PS00016	112->115	RGD	PDOC00016
PS00430	1->22	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKFZphfbr2_64j18.1)

DKFZphfbr2_64k24

group: transmembrane proteins

DKFZphfbr2_64k24 encodes a novel 412 amino acid protein with weak similarity to several known proteins.

The novel protein contains 5 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to AMAC1 "testicular condensing enzyme" ;
membrane regions: 5
Summary DKFZphfbr2_64k24 encodes a novel 412 amino acid protein, with
similarity to AMAC1; product: "testicular condensing enzyme"

similarity to AMAC1 "testicular condensing enzyme"

complete cDNA, complete cds, EST hits

Sequenced by Qiagen

Locus: unknown

Insert length: 1958 bp

Poly A stretch at pos. 1939, polyadenylation signal at pos. 1918

```
1 GGGCCCGCCT CGATTTTCCC AGGCGAGGGC ACGCCCGCGT CAGTCGCCTC
51 CGGGGACACCT TCCTCGCCAC GACACGCGAG TAACCGGGCC CCGGGAGCCG
101 GTCGGCGGCG GCGGACTGGG ACCTTGATCC TGCCCGCCCG GCCGCCCGAC
151 AAGGGAATGA GAGCGGACCC CGAACTCCAC ACACCCGCGT TTAGCCGCCA
201 CACCTAAGGG GCAGAACAGT CTTTGTGGGT AAGGGCCGGG CTGGGGCGGA
251 CGCGCCCCCG CCGCTTTGCA GACTTCGGGG TGCTCTGCAC GACGCCTGAA
301 AGGCCGCGGG GCCCGCATTT CTCTGTGCTG CCCTCCTGGA GAACCGGGAC
351 ACGGGGACGG GAGGGCCAGC ATCGGCTAGG GCCCGGTTTC CCGTTTCTTT
401 CCTCTGTGCG GTCTGGGCCC TCCTGCAGCG TCCATGATGA AGGCCAGGGG
451 CTGTTGCTTT CCTCTCGCCC AGTAGCCAAC CCAAGCAAGG GAATTAATTA
501 TCTGAAGAAA TGGATACTTC TCCCTCCAGA AAATATCCAG TTAATAAACG
551 GGTGAAAATA CATCCCAACA CAGTGATGGT GAAATATACT TCTCATATC
601 CCCAGCCTGG CGATGATGGA TATGAAGAAA TCAATGAAGG CTATGGGAAT
651 TTTATGGAGG AAAATCCAAA GAAAGGTCTG CTGAGTGAAA TGAATAAAAA
701 AGGGAGAGCT TTCTTTGGAA CCATGGATAC CCTACCTCCA CCAACAGAGG
751 ACCCAATGAT CAATGAGATT GGACAATTCC AGAGCTTTGC AGAAAAAAC
801 ATTTTTCAT CCAGAAAAAT GTGGATAGTG CTGTTTGGAT CTGCTTTGGC
851 TCATGGATGT GTAGCTCTTA TCACTAGGCT TGTTCCTGAT CGGTCTAAAG
901 TTCCATCTCT AGAAGTCTAT TTTATCCGTT CTGTTTTCCT GGTCTTATCT
951 GTGTTAGTTG TGTGTTACTA TCAGGAGGCC CCGTTTGGAC CCAGTGGATA
1001 CAGATTACGA CTCTTCTTTT ATGGTGTATG CAATGTCAAT TCTATCACTT
1051 GTGCTTATAC ATCATTTTCA ATAGTTCCTC CCAGCAATGG GACCACTATG
1101 TGGAGAGCCA CAATACAGT CTTCACTGCC ATTTTGGCTT TTTTACTCGT
1151 AGATGAGAAA ATGGCTTATG TTGACATGGC TACAGTTGTT TGCAGCATCT
1201 TAGGTGTTTG TCTTGTCTAT ATCCCAAAAC TTGTTGATGA AGACAATTCT
1251 TTGTTAAATG CCTGGAAGA AGCCTTTGGG TACACCATGA CTGTGATGGC
1301 TGGACTGACC ACTGCTCTCT CAATGATAGT ATACAGATCC ATCAAGGAGA
1351 AGATCAGCAT GTGACTGCG CTGTTTACTT TTGGTTGGAC TGGGACAATT
1401 TGGGGAATAT CTAATATGTT TATTCTTCAA GAACCCATCA TCCATTAGA
1451 TGGAGAAACC TGGAGTTATC TCATTGCTAT ATGTGCTCTG TCTACTGCAG
1501 CATCTTAGG AGTTTATTAT GCCTTGGACA AATTCCATCC AGCTTTGGTT
1551 AGCACAGTAC AACATTGGA GATTGTGGTA GCTATGGTCT TGCAGCTTCT
1601 CGTGCTGCAC ATATTTCCCTA GCATCTATGA TGTTTTTGA GGGGTAATCA
1651 TTATGATTAG TGTTTTTGTG CTGCTGGGCT ATAACTTTA CTGGAGGAAT
1701 TTAAGAAGGC AGGACTACCA GGAAATACTA GACTCTCCCA TTAATGAAT
1751 ACCTGATTAT TATTGTCTCA TTAATGTTCA GTTATTAAATA TGTATACTGC
1801 CATTTTAAATG TTTACCTATG AATGTCTTTT GTGTTATATA ACTGACAGAG
1851 TGCTATAAAA TATATAATAT ATACAAATGC AGAAAAATTA TTCTAGTCTA
1901 ATATATTCAA ATACAAATAT TAAATATATG AAATACGTGA AAAAAAATA
1951 AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 510 bp to 1745 bp; peptide length: 412
Category: similarity to known protein

```
1 MDTSPSRKYP VKKRVKIHPN TVMVKYTSY PQPGDDGYEE INEGYGNFME
51 ENPKKGLLSE MKKKGRAFFG TMDTLPPPT DPINEIGQF QSFAEKNIFQ
101 SRKMMIVLFG SALAHGCVAL ITRLVSDRSK VPSLELIFIR SVFQVLSVLV
151 VCYYQEAFFG PSGYRLRLFF YGVCNVISIT CAYTSFSIVP PSNGTTMWRA
201 TTTVFSAILA FLLVDEKMAV VDMATVVCSI LGVCLVMIPN IVDENSLN
251 AWKEAFGYTM TVMAGLTAL SMIVYRSIKE KISMWTALFT FGWTGTIWI
301 STMFLQEP IPLDGETWSY LIAICVCSTA AFLGVYALD KFHPALVSTV
351 QHLEIVVAMV LQLLVLIHIFP SIYDVFGGVI IMISVFVLAG YKLYWRNLRR
401 QDYQEILDSP IK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_64k24, frame 3

TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds., N = 1, Score = 191, P = 1.9e-12

TREMBL:BMAJ733_6 product: "hypothetical protein"; Bacillus megaterium bgaM gene, N = 1, Score = 137, P = 1.6e-06

PIR:G71841 hypothetical protein jhpl155 - Helicobacter pylori (strain J99), N = 1, Score = 129, P = 1.3e-05

>TREMBLNEW:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus musculus testicular condensing enzyme (AMAC1) mRNA, complete cds.
Length = 362

HSPs:

Score = 191 (28.7 bits), Expect = 1.9e-12, P = 1.9e-12
Identities = 39/105 (37%), Positives = 66/105 (62%)

```
Query:  289 FTFGWTGTIWGISTMFILQEPIIPLDGETWSYLIAICVCSTAFLGVYALDKFHPALVS 348
          F FG  G + + +F+LQ P++P D +WS ++A+ + + +F+ V YA+ K HPALV
Sbjct:  248 FLFGLVGLMVSVPGLFVLQTPVLPQDTLSWSCVAVGLLALVSFVCVSYAVTKAHPALVC 307

Query:  349 TVQHLEIVVAMVLQLLVLH--IFPSIYDVFGGVIIMISVFVLAGYKL 393
          V H E+VVA++LQ VL+ + PS D+ G +++ S+ ++ L
Sbjct:  308 AVLHSEVVVALMLQYYVLYETVAPS--DIMGAGVVLGSIAIITAQNL 352
```

Pedant information for DKFZphfbr2_64k24, frame 3

Report for DKFZphfbr2_64k24.3

```
[LENGTH]      412
[MW]           46449.87
[PI]           6.99
[HOMOL]        TREMBL:AF016712_1 gene: "AMAC1"; product: "testicular condensing enzyme"; Mus
musculus testicular condensing enzyme (AMAC1) mRNA, complete cds. 8e-14
[PROSITE]      MYRISTYL 6
[PROSITE]      CK2_PHOSPHO_SITE 3
[PROSITE]      PKC_PHOSPHO_SITE 4
[PROSITE]      ASN_GLYCOSYLATION 1
[KW]           TRANSMEMBRANE 5
```

SEQ MDTSPSRKYPVKKRVKIHPNTVMV KYTSY PQPGDDGYEE INEGYGNFMEENPKKGLLSE

```

PRD      cccccccccccccceeecccceeeeeeccccccccceeeeeeccccccccccccchhhh
MEM      .....

SEQ      MKKKGRAFFGTTMDTLPPPTEDPMINEIGQFSQAEKNIFQSRKMWIVLFGSALAHGCVAL
PRD      hhhhccceeeccccccccccccceeecccchhhhhhhhhccceeeeeeccccchhhhhc
MEM      .....

SEQ      ITRLVSDRSKVPSLELIFIRSVFQVLSVLVVCYYQEAFFGPSGYRLRLFYGVNCVISIT
PRD      chhhhccccccccchhhhhhhhhhhheeeeeeccccccccceeeeeeccccceeeeee
MEM      .....MMMMMMMMMMMMMMMM.....

SEQ      CAYTSFSIYPPSSNGTTMRWATTTVFSAILAFLVLDEKMAYVDMATVVCISILGVLVMIPI
PRD      eccccceeeccccccccceeeehhhhhhhhhhhhhhhheeeeeecccccccccccccccc
MEM      .....

SEQ      IVDEDNSLLNAWKEAFGYTMTVMAGLTTALSIMIVYRSIKERISMWTALTFWGTGTIWI
PRD      cccccchhhhhhhhhhhheeeeeeehhhhhhhchhhhhhhhhhhccccccccceeeeee
MEM      .....MMMMMMMMMMMMMMMMMMMM.....

SEQ      STMFILQEPIIPLDGETWSYLIAICVSTAALFVGYALDKFHPALVSTVQHLIIVAMV
PRD      ceeeeeeccccccccceeeeeecccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhh
MEM      MMMMMMMM.....MMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMM

SEQ      LQLLVLIHFPSIYDFVGGVIIMISVFLAGYKLYWRNLRRODYQEILDSPIK
PRD      hhhhhhhhhccccccccceeeeeeccccccccchhhhhhhhhhhhhhhhhcccccc
MEM      MMMMMMMM.....MMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMM

```

Prosite for DKFZphfbr2 64k24.3

PS00001	193->197	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	126->129	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00006	92->96	CK2_PHOSPHO_SITE	PDOC00006
PS00006	277->281	CK2_PHOSPHO_SITE	PDOC00006
PS00006	371->375	CK2_PHOSPHO_SITE	PDOC00006
PS00008	70->76	MYRISTYL	PDOC00008
PS00008	88->94	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	265->271	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2 64k24.3)

DKF2phfbr2_6a17

group: brain derived

DKF2phfbr2_6a17 encodes a novel 100 amino acid protein with very weak similarity to human finger protein zFOC1.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 1424 bp

Poly A stretch at pos. 1405, polyadenylation signal at pos. 1389

```
1 GGGACTGAGG GGGTGGGCTT ACTCCCTGGG CAGTCTTGGG GGCCAGAGCT
51 GAGGCCAGTC CATATTACAG TGGCTGGGCT GTTTTTTTCA GTAGCCCCTA
101 GCATTGGCTG GGATTCCTGT TCCTGGGTGC GCCTCCACCT CCCTTCTGAT
151 GCTTCCTGGC TATGGTGGGG TGGGAACCTC AGTTTCCCCC AAAGTCTTCC
201 CTGGATGCTG GCTTCAGGTT GAAGACCCCT GTTCTTCCAG TTCTTCACGG
251 CTAGGTAGG GGCTCCTGCA TCACCTTCAG AATCAGTTCC AACCCCACT
301 CTCTTAGGCT TTTGTGCTCT GCTCTGCCCT GCCAGGCTGC CCTTGTCCAT
351 GTGAGTAGCA TGGGCGGGTG GTGGGGACGG CAGTGGTGAT GAAGGGGGTG
401 CACCACAGGC CTCATGAAGC AGTTCCACCA TGGGCGGTGT GCTGGGGCGT
451 GGCCACCACA GAGCACATGG CTGTGTCTAG GCGCAAGCAC TTTAGCAGTA
501 TCTGTTTACA TGGCAAGGA TCAAGCCGAC TACCTGTGCT GTCTACTGGG
551 ACAGCAGTCT CCGAGCTACT CCGTACCTCC CTCTGCCAGG TCGTGGAGTT
601 AGGCCCCAGT CCCTACTTGT CACTGGTTCC CACTGTGCTC CTAAGTGTGC
651 AGCACCTGGG AGCTCTGGCC TGGGGCTGGA GGCCCTGCTA GGAGCTGCAG
701 TTGGAGGCCG TTCTGTGCCC AGCAGCGGTG AGCGGCTCCC ATGGGCCCTG
751 TGTCTGCAGG GAGCCAGGGC TCGGCGACAT GTGCTGTGAA ACTGGCACCC
801 ACCTGGCGTG CTGCTGCCGC CACTTGCTTC CTGCAGCACC TCCTACCCCTG
851 CTCCGTGTCC TCCCTCTCCC CGCGCCTGGC TCAGGAGTGC TGGAAAAGCT
901 CACGCCTCGG CCTGGGAGCC TGGCCTCTTG ATATACCTCG AGCTTCCCTT
951 GTGCTCCCCA GCCCCAGGAC CACTGGCCCC TTGGCCTGAG GGGCTGGGGG
1001 CCCCACGACC TGCAGCGTCG AGTCCGGGAG AGAGCCCGGA GCGGCGTGCC
1051 ATCTCGGCTG GGCCTTGCTG AGAGCCTCCG CCCTGGCTTT CTCCCTGTCT
1101 GGTTCAGTGT GCTCACGTTG GTGCTACACA GCTAGAATAG ATATATTAG
1151 AGAGAGAGAT ATTTTAAAGA CAAAGCCAC AATTAGCTGT CCTTAAACAC
1201 CCGAGAAGCC CCTCCAGAA GAAGAGCGAT CCCTCGGACG GTCCGGGCGG
1251 GCACCTCAG CCGGGCTCTT TGCAGAAACA GCACCGCTGA CTGTGGGCC
1301 GGCCCTCAGA TGTGTACATA TACGGCTATT TCCTATTTTA CTGTTCTTCA
1351 GATTTAGTAC TTGTAATAAA ACACACACAT TAAGGAGAGA TTAACATT
1401 TTGCCAAAAA AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 389 bp to 688 bp; peptide length: 100

Category: putative protein

```
1 MKGVHHRPHE AVPTWACGWG VATTEHMAVS RRRKHFSSICL HAQGSRLPV
51 LSTGTAVSEL LRSLCQVVE LGPSPYLSLV PTVLLTVQHL GALANGWRPW
```

BLASTP hits

Entry S70007 from database PIR:
finger protein zfOCl - human (fragment)
Length = 183
Score = 62 (21.8 bits), Expect = 0.24, Sum P(2) = 0.22
Identities = 18/47 (38%), Positives = 24/47 (51%)

Alert BLASTP hits for DKFZphfbr2_6a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6a17, frame 2

Report for DKFZphfbr2_6a17.2

[LENGTH] 100
[MW] 10944.82
[pI] 9.49
[PROSITE] MYRISTYL 2
[PROSITE] PKC_PHOSPHO_SITE 2
[KW] Alpha_Beta

SEQ MKGVHHRPHEAVPTWACGWGVATTEHMAVSRKHFSSICLHAQGSSRLPVLSTGTAVSEL
PRD cccccccccccccccccccccchhhhhhhhhccccccccccccccccccccchhhhh

SEQ LRTSLCQVVELGSPYLSLVPTVLLTVQHLGALAWGWRPW
PRD hhhhheeeccccccccccccchhhhhhhhhchhhhhcccc

Prosites for DKFZphfbr2_6a17.2

PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	45->48	PKC_PHOSPHO_SITE	PDOC00005
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	54->60	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_6a17.2)

DKFZphfbr2_6b24

group: metabolism

DKFZphfkd2_6b24 encodes a novel 334 amino acid protein with similarity to several bacterial dTDP-4-dehydrorhamnose reductases (EC 1.1.1.133).

The novel protein seems to be a human enzyme similar to dTDP-4-dehydrorhamnose reductases. EC 1.1.1.133 catalyses the reaction: dTDP-6-deoxy-L-mannose + NADP(+) \rightleftharpoons dTDP-4-dehydro-6-deoxy-L-mannose + NADPH.

The new protein can find application in modulation of rhamnose metabolism and as a new enzyme for biotechnologic production processes.

similar to dTDP-6-deoxy-L-mannose-dehydrogenases

complete cDNA, EST hits, complete cds
Nucleotide sugars metabolism seems to be a dehydrogenase
localisation: region of primer A missing

Sequenced by AGOWA

Locus: /map="5"

Insert length: 2054 bp

Poly A stretch at pos. 2028, polyadenylation signal at pos. 2015

```
1 GGGGGAGGCC CGCGTCGATC CTGGGTTGGA GGAGGTGGCG GCCGCTGAGG
51 CTGCGGCGTG AAGACGGCGG GCATGGTGGG GCGGGAGAAA GAGCTCTCTA
101 TACACTTTGT TCCCGGGAGC TGTCGGCTGG TGGAGGAGGA AGTTAACATC
151 CCTAATAGGA GGGTTCTGGT TACTGGTGCC ACTGGGCTTC TTGGCAGAGC
201 TGTACACAAA GAATTTTCAGC AGAATAATTG GCATGCAGTT GGCTGTGGTT
251 TCAGAGAGAGC AAGACCAAAA TTGAACAGG TTAATCTGTT GGATTCATAT
301 GCAGTTTCATC ACATCATTCA TGATTTTCAG CCCCATGTTA TAGTACATTG
351 TGCAGCAGAG AGAAGACCAG ATGTTGTAGA AAATCAGCCA GATGCTGCCT
401 CTCACCTTAA TGTGGATGCT TCTGGGAATT TAGCAAAGGA AGCAGCTGCT
451 GTTGGAGCAT TTCTCATCTA CATTAGCTCA GATTATGTAT TTGATGGAAC
501 AAATCCACCT TACAGAGAGG AAGACATACC AGCTCCCTTA AATTGTATG
551 GCAAAACAAA ATTAGATGGA GAAAAGGCTG TCCTGGAGAA CAATCTAGGA
601 GCTGCTGTTT TGAGGATTCC TATTCTGTAT GGGGAAGTTG AAAAGCTCGA
651 AGAAAGTGCA GTGACTGTTA TGTTTGATAA AGTGCAGTTC AGCAACAAGT
701 CAGCAAAACAT GGATCACTGG CAGCAGAGGT TCCCCACACA TGTCAAAGAT
751 GTGGCCACTG TGTGCCGCGA GCTAGCAGAG AAGAGAATGC TGGATCCATC
801 AATTAAAGGA ACCTTTCAC TGTCTGGCAA TGAACAGATG ACTAAGTATG
851 AAATGGCATG TGCAATTGCA GATGCCTTCA ACCTCCCGAG CAGTCACTTA
901 AGACCTATTA CTGACAGCCC TGCTCTAGGA GCACAACGTC CGAGAAATGC
951 TCAGCTTGAC TGCTCCAAAT TCGAGACCTT GGGCATTGGC CAACGAACAC
1001 CATTTCGAAT TGAATCAAA GAATCACTTT GCCCTTTCCT CATGACAAAG
1051 AGATGGAGAC AAACGGCTCT TCATTAGTTT ATTTGTGTTG GGTTCCTTTT
1101 TTTTAAAT GAAAAGTATA GTATGTGGCC CTTTTAAAG AACAAAGGAA
1151 ATAGTTTGT ATGAGTACTT TAATTGTGAC TCTTAGGATC TTTAGGTAA
1201 ATGATGCTCT TGCCTAGTG AAATTGTCTA AAGAACTAA AGGGCAGTCA
1251 TGCCCTGTTT GCAGTAATTT TTCTTTTAT CATTATGTTT GTCCCTGGCTA
1301 AACTTGGAGT TTGAGTATAG TAAATTATGA TCCTTAAATA TTTGAGGGTC
1351 AGGATGAAGC AGATCTGCTG TAGACTTTTC AGATGAAAT GTTCATTCTC
1401 GTAACCTCCA TATTTTCAGG ATTTTGAAG CTGTTGACCA TTTGATGTTG
1451 ATTATTTTAA ATTGTGTGGA ATAGTATAAA AATCATTGGT GTTCATTATT
1501 TCCTTTGCC T GAGCTCAGAT CAAAATGTTT GAAGAAAGGA ACTTATTTT
1551 TGCAAGTTAC GTACAGTTT TATGCTTGAG ATATTTCAAC ATGTTATGTA
1601 TATTGGAAC TCTACAGCTT GATGCCTCCT GCTTTTATAG CAGTTTATGG
1651 GGAGCACTTG AAAGAGCGTG TGTACATGTA TTTTCTTCT AGGCAACAT
1701 TGAATGCAAA CGTGATTTT TTTAATATAA ATATATAACT GTCCTTTTCA
1751 TCCCATGTTG CCGCTAAGTG ATATTTCATA TGTGTGGTTA TACTCATAAT
1801 AATGGGCTT GTAAGTCTT TCACCATTCA TGAATAATAA TAAATATGTA
1851 CTGCTGGCAT GTAATGCTTA GTTTTCTTGT ATTTACTTCT TTTTAAATAA
1901 TGAAGGACC AAATCTCTAA ACTAATTGTT CTTTGTGTC TTTAATTTT
1951 AAAAATTACA TTCTTCTGAT GTAACATGTG ATACATACAA AAGAATATAG
2001 TTTAATATGT ATTGAAATAA AACACAATAA AATTAAAAAA AAAAAAATAA
2051 AAAA
```

BLAST Results

Entry G37115 from database EMBL:
SHGC-56899 Human.Homo sapiens STS genomic.
Score = 446, P = 4.6e-14, identities = 90/91

Medline entries

99109950:
The metabolism of 6-deoxyhexoses in bacterial and animal cells.

Peptide information for frame 1

ORF from 73 bp to 1074 bp; peptide length: 334
Category: similarity to known protein

1 MVGREKELSI HFVPGSCRLV EEEVNIPNRR VLVGTATGLL GRAVHKEFQO
51 NNWHA VGCGF RRARPKEQV NLLDSNAVHH IIHDFQPHVI VHCAAERRPD
101 VVENQPDAA S QLNVDASGNL AKEAAVGAFLIYISSDYVFDG-TNPPYR 148
151 DIPAPLNLYG KTKLDGEKAV LENNLGA AVL RIPILYGEVEKLEESAVTVMF 208
201 FOKVQFSNKS ANMDHWQQRFP THVKDVATV CRQLAEKRLDPSIKGT 267
251 SGNEQMTKYE MACAIADAFN LPSSHLRPIT DSPVLGAQRP RNAQLDCSKL 305
301 ETLGIGQRTF FRIGIKESLW PFLIDKRWRQ TVFH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6b24, frame 1

PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans, N = 1, Score = 293, P = 6.4e-26

TREMBL:SSU51197_21 gene: "rhsD"; product:
"dTDP-6-deoxy-L-mannose-dehydrogenase"; Sphingomonas S88 sphingan
polysaccharide synthesis (spsG), (spsS), (spsR), glycosyl transferase
(spsQ), (spsI), glycosyl transferase (spsK), glycosyl transferase
(spsL), (spsJ), (spsF), (spsD), (spsC), (spsE), Urf 32, Urf 26,
ATP-binding cassette trans., N = 1, Score = 291, P = 1e-25

SWISSPROT:RFB_DHISN PROBABLE DTDP-4-DEHYDRORHAMNOSE REDUCTASE (EC
1.1.1.133) (DTDP-4-KETO- L-RHAMNOSE REDUCTASE) (DTDP-6-DEOXY-L-MANNOSE
DEHYDROGENASE) (DTDP-L- RHAMNOSE SYNTHETASE), N = 1, Score = 283, P =
7.4e-25

>PIR:T00104 probable dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133) -
Actinobacillus actinomycetemcomitans
Length = 294

HSPs:

Score = 293 (44.0 bits), Expect = 6.4e-26, P = 6.4e-26
Identities = 89/276 (32%), Positives = 151/276 (54%)

Query: 30 RVLVTGATGLLGRAVHKEFQQNNWHA VGCGFRRARPKEQVNNLLDSNAVHHIIHDFQPHV 89
R+L+TGA G LGR++ K N + V F +++++ + + V II F+P+V
Sbjct: 3 RLLITGAGGQLGRSLAKLLVDNGRYEV-----LALDFSELDITNKDMVFSTIIDSFKPNV 56

Query: 90 IVHCAAERRPDVVENQPDAA S QLNVDASGNLAKAEAAVGAFLIYISSDYVFDG-TNPPYR 148
I++ AA D E + +A +NV LA+ A + +++S+DYVFDG + Y+
Sbjct: 57 IINAAAYTSVDQAELEVSSAYS VNVRGVQYLAEAAIRHNSAILHVSTDYVFDGYKSGKYK 116

Query: 149 EEDIPAPLNLYGKTKLDGEKAVLENNLGA AVL RIPILYGEVEKLEESAVTVMF 208
E DI PL +YK+K +GE+ +L + + +LR +GE + V M ++ +
Sbjct: 117 ETDIIHPLCVYGKSKAEGERLLLTSPKSIILRTSWTFGEYGN---NFVKTML-RLAKNR 172

Query: 209 KSA NMDHWQQRFP THVKDVATVCRQLAEKRLDPSIK-GTFHWSGNEQMTKYEMACAIAD 267
+ Q PT+ D+A+V Q+AEK ++ ++K G +H++G ++ Y+ A AI D
Sbjct: 173 DIIGVVADQIGGPTYSGDIASVLIQIAEKIIVGETVKYGIYHFTGEPVSWYDFAIAIFD 232

Query: 268 AF-----NLPSSHLRPITDSPVLGAQRP RNAQLDCSKLE-TLGI 305
N+P + D P L A+RP N+ LD +K++ GI
Sbjct: 233 EAVAQKVLENVPLVNAITADYPTL-AKRPANSCLDLTKIQQAFGI 277

Report for DKFZphfbr2_6b24.1

```
SEQ      MVGREKELSIHFVPGSCLRVEEVNI PNRRVLVTGATLLGRAVHKFEQQNNWHA VGGCF
PRD      cccccceeeccccceeeccccceeeccccceeeccccchhhhhhhhhhhccceeeeeecc

SEQ      RRARPKFEQVNLDSNAVHHIIHDFQPHVIVCEAAERRPDVVENQPDASQLNVDASGML
PRD      cccccccccccccchhhhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhhhhhhhccchh

SEQ      AKEAAAVGAFLIYISSDYVDGNTNPPYREDIPAPLNLYGKTKLDGEKAVLENNLGA AVL
PRD      hhhhhhhhhheeeeeccccccccccccccccccccccccccccchhhhhhhhhccccceeee

SEQ      RIPILYGEVEKLEESAVTMFMDKVQFSNKSANMDHWQQRFPPTHVKDVATVCRQLAEKRML
PRD      eeeeeccccccccchhhhhhhhhhhhhhhccceccccccccccccchhhhhhhhhhhhhhhhh

SEQ      DPSIKGTFWHSNEQMTKYEMACATADAFNLPSSHLRPITDSPVLGAQRPRNAQLDCSKL
PRD      cccccceeeccccccccchhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhh

SEQ      ETLGIGQRTFPRIGIKESLWPFLLDKRWQTVFH
PRD      hhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
```

PS000001	208->212	ASN GLYCOSYLATION	PDOC000001
PS000005	16->19	PKC_PHOSPHO_SITE	PDOC000005
PS000005	207->210	PKC_PHOSPHO_SITE	PDOC000005
PS000005	243->246	PKC_PHOSPHO_SITE	PDOC000005
PS000006	162->166	CK2_PHOSPHO_SITE	PDOC000006
PS000006	251->255	CK2_PHOSPHO_SITE	PDOC000006
PS000006	257->261	CK2_PHOSPHO_SITE	PDOC000006
PS000006	298->302	CK2_PHOSPHO_SITE	PDOC000006
PS000008	314->320	MYRISTYL	PDOC000008

298

DKFZphfbr2_6i20

group: brain derived

DKFZphfbr2_6i20 encodes a novel 296 amino acid protein with similarity to ribosomal protein L15 precursor of *S. cerevisiae* mitochondria.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to ribosomal protein L15 precursor, mitochondrial

complete cDNA, complete cds, EST hits
potential mitochondrial L15 ribosomal protein

Sequenced by AGOWA

Locus: /map="377.5 cR from top of Chr8 linkage group"

Insert length: 1122 bp

Poly A stretch at pos. 1099, polyadenylation signal at pos. 1071

```

1  GGGGGCCCTT  GAAAGTCTCT  GGATCTGCGG  GTTATGGCCG  GTCCTTGCA
51  GGGCGGTGGG  GCCCGGGCCC  TGGACCTACT  CCGGGGCCTG  CCGCGTGTGA
101 GCGTGGCCAA  CTTAAAGCCC  AATCCCGGCT  CCAAGAAACC  GGAGAGAAGA
151 CCAAGAGGTC  GGAGAAGAGG  TAGAAAATGT  GGCAGAGGCC  ATAAAGGAGA
201 AAGGCAAGA  GGAACCCGGC  CCCGCTTGGG  CTTGAGGGA  GGCCAGACTC
251 CATTTTACAT  CCGAATCCCA  AAATACGGGT  TTAACGAAGG  ACATAGTTTC
301 AGAGCCAGT  ATAAGCCTAT  GAGTCTCAAT  AGACTGCAGT  ATCTTATTGA
351 TTTGGGTCGT  GTTGATCCTA  GTCAACCTAT  TGACTTAACC  CAGCTGTGCA
401 ATGGGAGAGG  TGTGACCATC  CAGCCACTTA  AAAGGGATTA  TGATGCCAG
451 CTGGTTGAGG  AGGGTGCTGA  CACCTTTACG  GCAAAAGTTA  ATATTGAAGT
501 ACAGTTGGCT  TCAGAACTAG  CTATTGCTGC  CATTGAAAAA  AATGGTGGTG
551 TTGTTACTAC  AGCCTTCTAT  GATCCAAGAA  GTCTGGACAT  TGTATGCAAA
601 CCTGTTCCAT  TCTTTCTTCG  TGGACAACCC  ATTCCAAAAA  GAATGCTTCC
651 ACCAGAAGAA  CTGGTACCAT  ATTACACTGA  TGCAAGAAGC  CGTGGGTACC
701 TGGCGGATCC  TGCCAAATTT  CCTGAAGCAC  GACTTGAACT  CGCCAGGAAG
751 TATGGTTATA  TCTTACCTGA  TATCACTAAA  GATGAACCTC  TCAAAATGCT
801 CTGTACTAGG  AAGGATCCAA  GGCAGATTTT  CTTTGGTCTT  GCTCCAGGAT
851 GGGTGGTGAA  TATGGCCGAT  AAGAAAATCC  TAAACCTTAC  AGATGAAAAA
901 CTCCTTAAGT  ATTATACCTC  ATGAATTCCT  GTCCAGGAA  GCAGAGTTGT
951 TAAAGAGTAC  TGAATAGGG  GCTGAAGGAT  CTATATTCCT  TTATTGCATT
1001 TTCCCTATGT  ATAATTTTCC  AGATGGTGAT  GTTACTTTTC  AGTGTACTCA
1051 TATGTCTCAT  TTTTCATCTA  AATTAAATGG  CAGGAAACAA  GGACTGCATA
1101 GAGAAAAAAA  AAAAAAAAAA  AA
```

BLAST Results

Entry HS500354 from database EMBL:

human STS WI-12392.

Length = 426

Minus Strand HSPs:

Score = 1791 (268.7 bits), Expect = 1.1e-74, P = 1.1e-74

Identities = 375/384 (97%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 34 bp to 921 bp; peptide length: 296

Category: strong similarity to known protein

1 MAGPLQGGGA RALDLLRGLP RVSLANLKP NPGSKKPERRP RGRRRGRKCG

51 RGHKGERQRG TRPRLGFEGG QTPFYIRIPK YGFNEGHSFR RQYKPSMLNR
 101 LQYLIDLGRV DPSQPIDLTQ LVNDRGVTIQ PLKRDYDVQL VEEGADTFTA
 151 KVNIEVQLAS ELAIAAIEKN GGVVTTAFYD PRSLDIVCKP VPFFLRGQPI
 201 PKRMLPPEEL VPYYTDAKNR GYLADPAKFP EARLELARKY GYLDPITKD
 251 ELFKMLCTRK DPRQIFFGLA PGWVVMADK KILKPTDENL LKYYTS

BLASTP hits

Entry S63258 from database PIR:
 ribosomal protein L15 precursor, mitochondrial - yeast (*Saccharomyces cerevisiae*)
 Length = 322
 Score = 259 (91.2 bits), Expect = 2.0e-22, P = 2.0e-22
 Identities = 71/200 (35%), Positives = 106/200 (53%)

Entry H70161 from database PIR:
 ribosomal protein L15 (rplO) - Lyme disease spirochete
 Length = 145
 Score = 173 (60.9 bits), Expect = 4.8e-13, P = 4.8e-13
 Identities = 45/140 (32%), Positives = 73/140 (52%)

Alert BLASTP hits for DKFZphfbr2_6i20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_6i20, frame 1

Report for DKFZphfbr2_6i20.1

[LENGTH] 296
 [MW] 33495.98
 [pI] 9.98
 [HOMOL] TREMBL:AF067212_1 gene: "F37F2.1"; *Caenorhabditis elegans* cosmid F37F2. 1e-38

[FUNCAT] 05.01 ribosomal proteins [S. cerevisiae, YNL284c] 7e-15
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YNL284c] 7e-15
 [FUNCAT] j mna translation and ribosome biogenesis [M. genitalium, MG169] 1e-06
 [BLOCKS] BL00475D
 [BLOCKS] BL00475B Ribosomal protein L15 proteins
 [PIRKW] ribosome 2e-13
 [PIRKW] mitochondrion 2e-13
 [PIRKW] protein biosynthesis 2e-13
 [SUPFAM] Escherichia coli ribosomal protein L15 4e-06
 [PROSITE] MYRISTYL 3
 [PROSITE] AMIDATION 2
 [PROSITE] CK2_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 4
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 12.50 %

SEQ MAGPLQGGGARALDLLRGLPRVSLANLKPNGSKKPERRPRGRRGRKCGRGHKGGERQRG
 SEGXXX
 PRD ccc

SEQ TRPRLGFEGGQTPFYIRIPKYGFNEGHSFRQYKPSMLNRLQYLIDLGRVDPSQPIDLTQ
 SEG
 PRD ccc

SEQ LVNDRGVTIQPLKRDYDVQLVEEGADTFTAKVNIEVQLASELAIAAIEKNGGVVTTAFYD
 SEG
 PRD ccc

SEQ PRSLDIVCKPVPFFLRGQPIPKRMLPPEELVPYYTDAKNRGYLADPAKFPPEARLELARKY
 SEG
 PRD ccc

SEQ GYLDPITKDLEFKMLCTRKDPRQIFFGLAPGWVVMADKKILKPTDENLLKYYTS
 SEG
 PRD ccc

Prosite for DKFZphfbr2_6i20.1

PS00005 33->36 PKC_PHOSPHO_SITE PDOC00005
 PS00005 88->91 PKC_PHOSPHO_SITE PDOC00005

PS00005	149->152	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00006	248->252	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	171->177	MYRISTYL	PDOC00008
PS00008	268->274	MYRISTYL	PDOC00008
PS00009	41->45	AMIDATION	PDOC00009
PS00009	45->49	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_6i20.1)

DKF2phfbr2_6017

group: nucleic acid management

DKF2phfbr2_6017 encodes a novel 455 amino acid protein with strong similarity to DEAD-box ATP-dependent RNA helicases YHR065c and T26G10.1.

The *S. cerevisiae* protein YHR065c is required for maturation of the 35S RNA primary transcript.

The new protein can find application in modulating rRNA maturation.

strong similar to RNA helicases

complete cDNA, complete cds, EST hits

probable start at Bp 27 matches kozak consensus ANNatgG

involved in maturation of r-RNA ??

YHR065c/Rrp3p is involved in maturation of the 35S primary transcript

Drs1p cold-sensitive mutation has slow 27S to 25S pre-rRNA

conversion and is deficient in 60S ribosomal subunits

Sequenced by AGOWA

Locus: unknown

Insert length: 1840 bp

Poly A stretch at pos. 1815, polyadenylation signal at pos. 1793

```
1 GGGGACTTCC GGAGACCTCA CACAAGATGG CGGCACCCGA GGAACACGAT
51 TCTCCGACCG AAGCGTCCCA GCCGATTGTG GAAGAGGAGG AAACATAAAC
101 ATTTAAAGAC CTGGGTGTGA CAGATGTGTT GTGTGAAGCT TGTGACCACT
151 TGGGATGGAC AAAACCCACC AAGATTGAGA TTGAAGCTAT TCCTTTGGCC
201 TTACAAGGTC GTGATATCAT TGGGCTTGCA GAAACTGGCT CTGGAAAGAC
251 AGGCGCCTTT GCTTTGCCCA TTCTAAACGC ACTGCTGGAG ACCCCGCAGC
301 GTTTGTTTGC CCTAGTTCTT ACCCCGACTC GGGAGCTGGC CTTTCAGATC
351 TCAGAGCAGT TTGAAGCCCT GGGGTCTCTT ATTGGAGTGC AGAGTGCTGT
401 GATTGTAGGT GGAATTGATT CAATGTCTCA ATCTTTGGCC CTGCAAAAAA
451 AACCACATAT AATAATAGCA ACTCCTGGTC GACTGATTGA CCACTTGGAA
501 AATACGAAAG GTTCAACTT GAGAGCTCTC AAATACTTGG TCATGGATGA
551 AGCCGACCGA ATACTGAATA TGGATTTTGA GACAGAGGTT GACAAGATCC
601 TCAAAGTGAT TCCTCGAGAT CGGAAAACAT TCCTCTTCTC TGCCACCATG
651 ACCAAGAAGG TTCAAAAACT TCAGCGAGCA GCTCTGAAGA ATCCTGTGAA
701 ATGTGCCGTT TCCTCTAAAT ACCAGACAGT TGAATAATTA CAGCAATATT
751 ATATTTTTAT TCCCTCTAAA TTCAAGGATA CCTACCTGGT TTATATTCTA
801 AATGAATTGG CTGGAAACTC CTTTATGATA TTCTGCAGCA CCTGTAATAA
851 TACCCAGAGA ACAGCTTTGC TACTGCGAAA TCTTGGCTTC ACTGCCATCC
901 CCCTCCATGG ACAAAATGAGT CAGAGTAAGC GCCTAGGATC CCTTAATAAG
951 TTTAAGGCCA AGGCCCGTTC CATTTCTCTA GCAACTGACG TTGCCAGCCG
1001 AGGTTTGGAC ATACCTCATG TAGATGTGGT TGTCAACTTT GACATTCCTA
1051 CCCATTCCAA GGATTACATC CATCGAGTAG GTCGAACAGC TAGAGCTGGG
1101 CGCTCCGGAA AGGCTATTAC TTTTGTGACA CAGTATGATG TGGAATCTTT
1151 CCAGCGCATA GAACACTTAA TTGGGAAGAA ACTACCAGGT TTTCCAACAC
1201 AGGATGATGA GGTATGATG CTGACAGAAC GCGTCGCTGA AGCCCAAGAG
1251 TTTGCCCGAA TGGAGTTAAG GGAGCATGGA GAAAAGAAGA AACGCTCGCG
1301 AGAGGATGCT GGAGATAATG ATGACACAGA GGGTGCTATT GGTGTCAGGA
1351 ACAAGGTGGC TGGAGGAAAA ATGAAGAAGC GGAAAGGCCG TTAATCACTT
1401 TTATGAAGGC TCGAGTTCTG CTGTTCTGTA AAAGAAAATT GGAGAATGAA
1451 ACCTGCTCCA ACAGAGATCA TGAGACTGAA ATTGGTCAGA ATTGTGTCCA
1501 GAATGTGCTC AGCTAATTCA GTATTCTTCC CCATTCTGGG TTGGAGTTTA
1551 CTGCAGAGTA ATTCTTACAG TGCTGATGTC AAGACTGTGA CTGTTCTTCG
1601 ACTTTGATTC CTTGCTCATG ACATGAGTAG GGTGTGCTCT TCTGTCACCT
1651 CACACAGACC TTTTGCCCTT TTTAGCTGCA AGTCAAGGAC TAGGTTGATG
1701 ATGCCCATGA CCTGTAATG TAAAGAAGCT TGGACATCTG CAAATGATAT
1751 TTAACCATC TTGGCTTGTG CTTTATTCAA ACTAATGTGA AACAAATAAT
1801 TTAATATTA TTTTAAAAAG AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 27 bp to 1391 bp; peptide length: 455
 Category: strong similarity to known protein

```

1 MAAPEEHDSP TEASQPIVEE EETKTFKDLG VTDVLCEACD QLGWTKPTKI
51 QIEAIPALQ GRDIIGLAET GSGKTGAFAL PILNALLETP QRLFALVLT
101 TRELAFQISE QFEALGSSIG VQSAVIVGGI DSMSQSLALA KKPHEIIATP
151 GRLIDHLENT KGFNLRLALKY LVMDEADRIL NMDFETEVDK ILKVIPDRK
201 TFLFSATMTK KVQKLQRAAL KNPVKCAVSS KYQTVEKLQQ YYIFIPSKFK
251 DTYLVYILNE LAGNSFMIFC STCNNTQRTA LLLRNLGFTA IPLHGQMSQS
301 KRLGSLNKFK AKARSILLAT DVASRGLDIP HVDVVVNFDI PTHSKDYIHR
351 VGR TARAGRS GKAITFVTQY DVELFQRIEH LIGKKLPGFP TQDDEVMMMLT
401 ERVAEQRF RMELEHGEK KRSREDAGD NDDTEGAIGV RNKVAGGKMK
451 KKKGR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_6017, frame 3

PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*, N = 1, Score = 1497, P = 1.6e-153

PIR:S46713 hypothetical protein YHR065c - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 1154, P = 3.6e-117

TREMBL:ATH010462_1 gene: "RH10"; product: "RNA helicase"; *Arabidopsis thaliana* mRNA for DEAD box RNA helicase, RH10, N = 1, Score = 1122, P = 8.9e-114

TREMBL:AC002985_2 product: "R27090_2"; Human DNA from chromosome 19-specific cosmid R27090, genomic sequence, complete sequence., N = 1, Score = 950, P = 1.5e-95

>PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - *Caenorhabditis elegans*
 Length = 489

HSPs:

Score = 1497 (224.6 bits), Expect = 1.6e-153, P = 1.6e-153
 Identities = 283/442 (64%), Positives = 364/442 (82%)

```

Query:   19 EEEETKTFKDLGVTDVLCEACDQLGWTKPTKIQIEAIPALQGRDIIGLAETGSGKTGAF 78
          E+ + K+F +LGV+ LC+AC +LGW KP+KIQ A+P ALQG+D+IGLAETGSGKTGAF
Sbjct:   39 EDVKEKSFAELGVSQPLCDACQRLGWMKPSKIQAALPHALQKDVIGLAETGSGKTGAF 98

Query:   79 ALPILNALLETPQRLFALVLTPTRELAFQISEQFEALGSSIGVQSAVIVGGIDSMSQSLA 138
          A+P+L +LL+ PQ F LVLTPTRELAFQI +QFEALGS IG+ +AVIVGG+D +Q++A
Sbjct:   99 AIPVLQSLLDHPQAFCLVLTPTRELAFQIQQFEALGSGIGLIAAVIVGGVDMAAQAMA 158

Query:   139 LAKKPHEIIATPGRLIDHLENTKGFNLRLALKYLVMDADRILNMDFETEVDKILKVIPRD 198
          LA+++PHII+ATPGRL+DHLENTKGFNL+ALK+L+MDEADRILNMDFE E+DKILKVIPR+
Sbjct:   159 LARRPHIIVATPGRLVDHLENTKGFNLKALKFLIMDEADRILNMDFEVELDKILKVIPRE 218

Query:   199 RKTFLFSATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQYYIFIPSKFKDITYLVYIL 258
          R+T+LFSATMTKKV KL+RA+L++P + +VSS+Y+TV+ L+Q+YIF+P+K+K+TYLVY+L
Sbjct:   219 RRTYLFSAATMTKKVSKLERASLRDPARVSVSSRYKTVDNLKQHYIFVPNKYKETYLVYLL 278

Query:   259 NELAGNSFMIFCSTCNNTQRTALLRNLGFTAIPPLHGQMSQSKRLGSLNKFKAKARSILL 318
          NE AGNS ++FC+TC T + A++LR LG A+PLHGQMSQ KRLGSLNKFK+KAR IL+
Sbjct:   279 NEHAGNSAIVFCATCATTMQIAVMLRQLGMQAVPLHGQMSQEKRLGSLNKFKSKAREILV 338

Query:   319 ATDVASRGLDIPHDVVVNFIDIPTHSKDYIHRVGR TARAGRS GKAITFVTQYDVELFQRI 378
          TDVA+RGLDIPHDV+V+N+D+P+ SKDY+HRVGR TARAGRS GAIT VTQYDVE +Q+I
Sbjct:   339 CTDVAARGLDIPHDVMVINYDMPQSKDYVHRVGR TARAGRS GIAITVVTQYDVEAYQKI 398

Query:   379 EHLIGKKLPGFPTQDDEVMMMLTERVAEQRFARMELREHGEKKK-----RSREDAGDND 433
          E +GKKL + ++EVM+L ER EA AR+E++E EKKK R +D GD ++
Sbjct:   399 EANLGKKLDEYKCVENEVMVLVERTQEATENARIEMKEMDEKKKSGKKRRQNDDFGDTTEE 458

Query:   434 TEGAIGVRNKVAGGKMKKKRGR 455

```

+ G + K GG+ GR
 Sbjct: 459 SGGRFKMGIKSMGGRGGSGGGR 480

Pedant information for DKFZphfbr2_6ol7, frame 3

Report for DKFZphfbr2_6ol7.3

[LENGTH] 455
 [MW] 50646.80
 [pI] 9.18
 [HOMOL] PIR:S40731 ATP-dependent RNA helicase homolog T26G10.1 - Caenorhabditis elegans
 1e-167
 [FUNCAT] 04.01.04 rRNA processing [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YHR065c] 1e-127
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YHR169w] 2e-79
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 1e-71
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-66
 [FUNCAT] j mRNA translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-63
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 1e-58
 [FUNCAT] 04.05.03 mRNA processing (splicing) [S. cerevisiae, YDL084w] 1e-55
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
 YOR204w] 5e-55
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 5e-55
 [FUNCAT] 1 genome replication, transcription, recombination and repair [H.
 influenzae, HI0892] 9e-48
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 2e-45
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-42
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 7e-16
 [FUNCAT] 03.19 recombination and DNA repair [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 7e-12
 [FUNCAT] r general function prediction [M. jannaschii, MJ1401] 5e-06
 [BLOCKS] BL00175B Phosphoglycerate mutase family phosphohistidine proteins
 [BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
 [BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
 [PIRKW] nucleus 4e-60
 [PIRKW] RNA binding 7e-69
 [PIRKW] DEAD box 7e-69
 [PIRKW] transmembrane protein 9e-41
 [PIRKW] DNA binding 3e-55
 [PIRKW] recF recombination pathway 3e-11
 [PIRKW] ATP 1e-126
 [PIRKW] purine nucleotide binding 7e-69
 [PIRKW] P-loop 1e-126
 [PIRKW] hydrolase 1e-55
 [PIRKW] protein biosynthesis 7e-69
 [PIRKW] ATP binding 3e-61
 [SUPFAM] ATP-dependent RNA helicase eIF-4A 8e-06
 [SUPFAM] WW repeat homology 4e-58
 [SUPFAM] translation initiation factor eIF-4A 7e-69
 [SUPFAM] DEAD/H box helicase homology 1e-126
 [SUPFAM] recQ helicase homology 5e-12
 [SUPFAM] ATP-dependent RNA helicase homology 8e-06
 [SUPFAM] unassigned DEAD/H box helicases 1e-126
 [SUPFAM] ATP-dependent RNA helicase DBP1 4e-60
 [SUPFAM] ATP-dependent RNA helicase DHH1 1e-58
 [SUPFAM] recQ protein 3e-11
 [SUPFAM] tobacco ATP-dependent RNA helicase DB10 4e-58
 [SUPFAM] Bloom's syndrome helicase 5e-12
 [PROSITE] DEAD_ATP_HELICASE 1
 [PROSITE] ATP_GTP_A 1
 [PROSITE] MYRISTYL 5
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] PKC_PHOSPHO_SITE 9
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Helicases conserved C-terminal domain
 [PFAM] DEAD and DEAH box helicases
 [KW] Alpha_Beta

SEQ MAAPEEHDSPTEASQPIVEEETKTFKDLGVTDVLCACDQLGWTKPTKIQIEAIPALQ
 PRD cccccccccccccchhhhhhhhhhhccccchhhhhhhhhcccccccccccccccccc
 SEQ GRDIIGLAETGSGKTGAFALPILNALLETPQRLFALVLTPTRELAQISEQFEALGSSIG
 PRD cceeeeeeccccceehhhhhhhhhccccceeeeeeccccchhhhhhhhhhhhhhhhhhhcc

```

SEQ  VQSAVIVGGIDSMQSLALAKKPHIIATPGRLLIDHLENTKGFNLRAKYLVMDEADRIL
PRD  eeeeeeeccchhhhhhhhhccceeeeeeccccccccccccccccccccceehhhhhhhh

SEQ  NMDFETEVDKILKVIPRDRKTLFSAATMTKKVQKLQRAALKNPVKCAVSSKYQTVEKLQQ
PRD  hhccchhhhhhhhhccchhhhhhhccchhhhhhhhhhhccceeeeeeccccchhhh

SEQ  YYIFIPSKFKDTYLVYILNELAGNSFMIFCSTCNNTQRTALLRLNGFTAIPLHGQMSQS
PRD  hhhhhhhhhhhhhhhhhhhhhccceeeeeeccchhhhhhhhhhhccceeeccccchhh

SEQ  KRLGSLNKFKAARSILLATDVASRGLDIPHDVVVNFDPITHSKDYIHRVGR TARAGRS
PRD  hhhhhhhhhhhhhhhccchhhhhhhccccceeeeeecccccccccccccccccccc

SEQ  GKAITFVTQYDVELFQRIEHLIGKKLPGFPTQDDEVMMLETERVAEQR FARMELREHGEK
PRD  cceeeeeeccchhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KKRSREDAGDNDDEGAIGVRNKVAGGKMKRKRGR
PRD  hhhhhcccccccccccccccccccccccccccccccccccc

```

Prosites for DKFZphfbr2_6ol7.3

PS00001	274->278	ASN_GLYCOSYLATION	PDOC00001
PS00004	421->425	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	229->232	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	300->303	PKC_PHOSPHO_SITE	PDOC00005
PS00005	354->357	PKC_PHOSPHO_SITE	PDOC00005
PS00005	360->363	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	25->29	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	391->395	CK2_PHOSPHO_SITE	PDOC00006
PS00006	424->428	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	128->134	MYRISTYL	PDOC00008
PS00009	382->386	AMIDATION	PDOC00009
PS00017	68->76	ATP_GTP_A	PDOC00017
PS00039	172->181	DEAD_ATP_HELICASE	PDOC00039

Pfams for DKFZphfbr2_6ol7.3

HMM_NAME	DEAD and DEAH box helicases		
HMM	*gLPWILRnIyeMGFEkPTPIQQqAIPiILeGRDVMACQTSgSKTAAFG ++ ++++++G++KPT+IQ +AIP++L+GRD+++ A TGSGKT+AF		
Query	30	GVTDVLCEACDQLGWTkPTKIQIEAIPALQGRDIIGLAETGSGKTGAF	78
HMM	lIPMLQHIdwdPwppPQdPrALILAPTRELAMQIEEcRkFgkHMNgIR ++P+L ++++P + ++AL+L+PTRELA QI+E++++G++++ ++		
Query	79	ALPILNALLETp----QR-LFALVLTPTRELAfQISEQFEALGSSIG-VQ	122
HMM	ImcIYGGtnMRdQMRmLeRGpPHIVATPGRLLIDHIER.gtldLDrIeML +++I+GG + + Q L+++P HI+IATPGRLLIDH+E+ ++L++++L		
Query	123	SAVIVGGIDSMQSLALAKKP-HIIATPGRLLIDHLENTKGFNLRAKYL	171
HMM	VMDEADRMLDMGFIDQIRrImrqIPmpwNRQTMMFSATMPdeIqELARrF VMDEADR+L+M+F+ ++++I++ IP ++R T +FSATM++++Q+L+R+		
Query	172	VMDEADRILNMDFETEVDKILKVIP--RDRKTLFSAATMTKKVQKLQRAA	219
HMM	MRNPiRInIdMdElTtnEnIkQwYiyVerEMWkfcdLcrLle* ++NP+ ++ +++++T++ ++Q+YI+++ + K +L+++++		
Query	220	LKNPVKCAVSSKYQTVE-KLQQYYIFIP-SKFKDTYLVYILN	259

HMM_NAME Helicases conserved C-terminal domain

HMM *EileeWLknlGIrvmYIHGdMpQeERdeIMddFnNGEynVLicTDVggr

```
Query      277  ++ + L+NLG++++ +HG+M+Q +R+  +++F++  +L++TDV++R  325
              QRTALLLRNLGFTAIPLHGQMSQSKRLGSLNKFKAARSILLATDVASR
HMM          GIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
              G+DIP V++V+N+D+P ++ +YI+R+GRT+R+G
Query      326  GLDIPHDVVVNFDIPTHSKDYIHRVGRTARAG  358
```

DKFZphfbr2_71o20

group: brain derived

DKFZphfbr2_71o20 encodes a novel 232 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits
on genomic level encoded by AC006186 (3 exons)

Sequenced by GBF

Locus: /map="10q22.1"

Insert length: 1768 bp

Poly A stretch at pos. 1742, polyadenylation signal at pos. 1726

```
1 GGGGGCAGCA GGCCAAGGGG GAGGTGCGAG CGTGGACCTG GGACGGGTCT
51 GGGGGGCTCT CGGTGGTTGG CACGGGTTCT CACACCCATT CAAGCGGCAG
101 GACGCACCTG TCTTAGCAGT TCTCGCTGAC CGCGCTAGCT GCGGCTTCTA
151 CGCTCCGGCA CTCTGAGTTC ATCAGCAAAC GCCCTGGCGT CTGTCTCTAC
201 CATGCCTAGC CTTTGGGACC GCTTCTCTGTC GTCGTCCACC TCCTCTTCGC
251 CCTCGTCCTT GCGCCGAACT CCCACCCAG ATCGGCCGCC GCGCTCAGCC
301 TGGGGGTCCG CGACCCGGA GGAGGGGTTT GACCGCTCCA CGAGCCTGGA
351 GAGCTCGGAC TGCAGTCCC TGGACAGCAG CAACAGTGGC TTCGGGCCGG
401 AGGAAGACAC GGCTTACCTG GATGGGGTGT CGTTGCCCGA CTTCGAGCTG
451 CTCAGTGACC CTGAGGATGA AACTTGTGT GCCAACCTGA TGCAGCTGCT
501 GCAGGAGAGC CTGGCCAGG CGCGGCTGGG CTCTCGACGC CCTGCGCGCC
551 TGCTGATGCC TAGCCAGTTG GTAAGCCAGG TGGGCAAAGA ACTACTGCGC
601 CTGGCCTACA GCGAGCCGTG CGGCCTGCGG GGGGCGCTGC TGGACGTCTG
651 CGTGGAGCAG GGCAAGAGCT GCCACAGCGT GGGCCAGCTG GCACTCGACC
701 CCAGCCTGGT GCCACCTTC CAGCTGACCC TCGTGCTGCG CCTGGACTCA
751 CGACTCTGGC CCAAGATCCA GGGGCTGTTT AGCTCCGCCA ACTCTCCCTT
801 CCTCCCTGGC TTCAGCCAGT CCCTGACGCT GAGCACTGGC TTCCGAGTCA
851 TCAAGAAGAA GCTGTACAGC TCGGAACAGC TGCCCATGGA GGAGTGTGTA
901 ACTTCAACCT GAGGGGGCCG ACAGTGCCCT CCAAGACAGA GACGACTGAA
951 CTTTTGGGGT GGAGACTAGA GGCAGGAGCT GAGGGACTGA TTCCAGTGGT
1001 TGGAAAACCT AGGCAGCCAC CTAAAGTGGG GGTGGGGGAA TAGTGTCTCC
1051 CAGGAAGCTC ATTGAGTTGT GTGCGGGTGG CTGTGCATTG GGGACACATA
1101 CCCCTCAGTA CTGTAGCATG AAACAAAGGC TTAGGGGCCA ACAAGGCTTC
1151 CAGCTGGATG TGTGTGTAGC ATGTACCTTA TTATTTTGT TACTGACAGT
1201 TAACAGTGGT GTGACATCCA GAGAGCAGCT GGGCTGCTCC CGCCCCAGCC
1251 TGGCCCGAGG TGAAGGAAGA GGCACGTGCT CCTCAGAGCA GCCCGAGGGA
1301 AGGGGGAGGT CGGAGGTCGT GGAGGTGGTT TGTGTATCTT ACTGGTCTGA
1351 AGGGACCAAG TGTGTTTGTG GTTTGTTTGT TATCTTGTG TTCTGATCGG
1401 AGCATCACTA CTGACCTGTT GTAGGCAGCT ATCTTACAGA CGCATGAATG
1451 TAAGAGTAGG AAGGGGTGGG TGTCAGGGAT CACTTGGGAT CTTTGACACT
1501 TGAATAATTA CACCTGGCAG CTGCGTTTAA GCCTTCCCCC ATCGTGTACT
1551 GCAGAGTTGA GCTGGCAGGG GAGGGGCTGA GAGGGTGGG GCTGGAACCC
1601 CTTCCCGGGA GGAGTGCCAT CTGGGTCTTC CATCTAGAAC TGTTTACATG
1651 AAGATAAGAT ACTCACTGTT CATGAATACA CTTGATGTTT AAGTATTAAG
1701 ACCTATGCAA TATTTTCTAC TTTTCTAATA AACATGTTTG TTAAACAAA
1751 AAAAAAAAAA AAAAAAAA
```

BLAST Results

Entry AC006186 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 10 clone
CRI-JC2048 map 10q22.1; HTGS phase 1, 4 unordered pieces.

Score = 6512, P = 0.0e+00, identities = 1326/1345

3 exons

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 202 bp to 897 bp; peptide length: 232
 Category: putative protein

```

1 MPSSLWDRFSS SSTSSSPSSL PRTPTPDRPP RSAWGSATRE EGFDRSTSL
51 SSDCESLDSS NSGFGPEEDT AYLDGVSLPD FELLSDPEDE HLCANLMQLL
101 QESLAQARLG SRRPARLLMP SQLVSVQVKE LLRLAYSEPC GLRGALLDVC
151 VEQGSCHSV GQLALDPSLV PTFQLTLVLR LDSRLWPKIQ GLFSSANSFP
201 LPGFSQSLTL STGFRVIKKK LYSSEQLPIE EC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_71o20, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_71o20, frame 1

Report for DKFZphfbr2_71o20.1

```

[LENGTH]      232
[MW]           25354.60
[pI]           4.87
[PROSITE]      MYRISTYL      2
[PROSITE]      CK2_PHOSPHO_SITE      6
[PROSITE]      GLYCOSAMINOGLYCAN      1
[PROSITE]      PKC_PHOSPHO_SITE      1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      17.67 %

SEQ      MPSSLWDRFSSSSTSSSPSSLPRTPTPDRPPRSAWGSATREEGFDRSTSLSSDCESLDSS
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      NSGFGPEEDTAYLDGVSLPDPELLSDPEDEHLCANLMQLLQESLAQARLGSRPARLLMP
SEG      XX.....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      SQLVSVQVKELLRLAYSEPCGLRGALLDVCVEQGSCHSVGQLALDPSLVPTFQLTLVLR
SEG      .....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      LDSRLWPKIQGLFSSANSFPPLPGFSQSLTLSTGFRVIKKKLYSSEQLPIEEC
SEG      .....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

```

Prosite for DKFZphfbr2_71o20.1

PS00002	62->66	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	111->114	PKC_PHOSPHO_SITE	PDOC00005
PS00006	3->7	CK2_PHOSPHO_SITE	PDOC00006
PS00006	38->42	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS00008	141->147	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_71o20.1)

DKFZphfbr2_72b18

group: nucleic acid management

DKFZphfbr2_72b18 encodes a novel 715 amino acid protein with similarity to *E. coli* DNA-damage-inducible protein *dinP* and other proteins induced by DNA-damage.

The novel protein is similar to *dinP* of *E. coli*, *yqjH* of *B. subtilis*, *dinP* of *M. tuberculosis* and T19K24.15 of *A. thaliana*. The *dinB/P* pathway is a second SOS-pathway in *E. coli*. Therefore the new gene seems to be involved in DNA repair.

The new protein can find application in modulating DNA repair and mutagenesis.

similarity to DNA damage induced genes

complete cDNA, complete cds, potential start at Bp 49, EST hits
localisation primer site B is missing!

Sequenced by LMU

Locus: /map="416.0 cR from top of Chr18 linkage group"??

Insert length: 2475 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2431

```

1 GGGGGAGGAA GCGGGCGGCG ACGACGAGGA AGACGCCGAG GCCTGGGCCA
51 TGGAACTGGC GGACGTGGGG GCGGCAGCCA GCTCGCAGGG AGTTCATGAT
101 CAAGTGTTCG CCACACCAAA TGCTTCATCC AGAGTCATAG TACATGTGGA
151 TCTGGATTGC TTTTATGCAC AAGTAGAAAT GATCTCAAAT CCAGAGCTAA
201 AAGACAAACC TTTAGGGGTT CAACAGAAAT ATTTGGTGGT TACCTGCAAC
251 TATGAAGCTA GGAACTTGG AGTTAAGAAA CTTATGAATG TCAGAGATGC
301 AAAAGAAAAG TGTCCACAGT TGGTATTAGT TAATGGAGAA GACCTGACCC
351 GCTACAGAGA AATGTCTTAT AAGGTTACAG AATTACTGGA AGAATTTAGT
401 CCAGTTGTTG AGAGACTTGG ATTTGATGAA AATTTGTGG ATCTAACAGA
451 AATGGTTGAG AAGAGACTAC AGCAGCTGCA AAGTGATGAA CTTTCTGCGG
501 TGA CTGTGTC GGGTCATGTA TACAATAATC AGTCTATAAA CCTGCTTGAC
551 GTCTTGACAG TCAGACTACT TGTGGATCT CAGATTGCAG CAGAGATGCG
601 GGAAGCCATG TATAATCAGT TGGGGCTCAC TGGCTGTGCT GGAGTGGCTT
651 CTAATAAACT GTTGGCAAAA TTAGTTTCTG GTGTCTTTAA ACCAAATCAA
701 CAAACAGTCT TATTACCTGA AAGTTGTCAA CATCTTATTC ATAGTTTGAA
751 TCACATAAAG GAAATACCTG GTATTGGCTA TAAACTGCCC AAATGTCTTG
801 AAGCACTGGG TATCAATAGT GTGCGTGATC TCCAAACCTT TTCACCCAAA
851 ATTTTAGAAA AAGAATTAGG AATTTCAAGT GCTCAGCGTA TCCAAAAGCT
901 CAGTTTGGGA GAGGATAACT CCCCTGTGAT ACTCTCAGGA CCACCTCAGT
951 CCTTTAGTGA AGAAGATTCA TTTAAAAAAT GTACATCTGA AGTTGAAGCT
1001 AAAAATAAGA TTGAAGAACT ACTTGCTAGT CTTTAAACA GAGTATGCCA
1051 AGATGGAAGG AAGCCTCATA CAGTGAGATT AATAATCCGT CGGTATTCTT
1101 CTGAGAAGCA CTATGGTCGT GAGAGTCGTC AGTGCCCTAT TCCTTCACAT
1151 GTAATTCAGA AATTAGGGAC AGGAAATTAT GATGTGATGA CCCCATTGGT
1201 TGATATACCT ATGAAACTTT TCGAAATAT GGTGAATGTG AAGATGCCAT
1251 TTCACCTTAC CCTTCTAAGT GTGTGCTTCT GCAACCTTAA AGCACTAAAT
1301 ACTGCTAAGA AAGGGCTTAT TGATTATTAT TTAATGCCAT CATTATCAAC
1351 TACTTCACGC TCTGGCAAGC ACAGTTTAA AATGAAAGAC ACTCATATGG
1401 AAGATTTTCC CAAAGACAAA GAAACAAACC GGGATTTTCT ACCAAGTGGA
1451 AGAATTGAAA GTACAAGAAC TAGGGAGTCT CCACTAGATA CCACAAATTT
1501 TTCTAAAGAA AAAGACATTA ATGAATTCCT ACTCTGTTCA CTTCCTGAAG
1551 GTGTTGACCA AGAAGTCTCC AAGCAGCTTC CAGTAGATAT TCAAGAAAGAA
1601 ATCCTTCTG GAAAATCTAG GGAATAATTT CAAGGGAAG GAAGTGTGAG
1651 TTGTCCATTA CATGCCTCTA GAGGAGTATT ATCTTTCTTT TCTAAAAAAC
1701 AAATGCAAGA TATTCCATA AATCCTAGAG ATCATTATAT CAGTAGCAAA
1751 CAGGTATCCT CTGTATCTCC TTGTGAACCG GGAACATCAG GCTTTAATAG
1801 CAGTAGTTCT TCTTACATGT CTAGCCAAAA GGATTATTCA TATTATTTAG
1851 ATAATAGATT AAAAGATGAA CGAATAAGTC AAGGACCTAA AGAACCTCAA
1901 GGATTCCACT TTACAAATTC AAACCCTGCT GTGTCTGCTT TTCATTCAAT
1951 TCCAAACTTG CAGAGTGAGC AACTTTTCTC CAGAAACCAC ACTACAGATA
2001 GCCATAAGCA AACAGTAGCA ACAGACTCTC ATGAAGGACT TACAGAAAAT
2051 AGAGAGCCAG ATTCTGTTGA TGAGAAAATT ACTTTCCCTT CTGACATTGA
2101 TCCTCAAGTT TTCTATGAAC TACCAGAAGC AGTACAAAAG GAACTGCTGG
2151 CAGAGTGGAA GAGAACAGGA TCAGATTTCC ACATTGGACA TAAATAAGCA
2201 TATTCAAGCA AAAGGTCTGA AAAGCAAGGG AATACCATTA TTTTCGGATT
2251 AGCGGTTTAT TAAGCTCTTC TATATTAAAC ACTAATAGAT ATTCATAAAC
2301 GGAGTAAACT GTTCCAGATA AAGCAAGAA AGTTGCAAGA AGTAAATCTT
2351 GGCACAAAGC GTAAAAATAT AACAGAAGAA ATAATGTAAA ATACTATCTT
2401 TTATGTCTAA AGCCATTTTA TATTACTTTT CAATAAAAAG AATATCATGG
2451 TCAAAAAAAA AAAAAAAA AAAAC

```

BLAST Results

 Entry HS086339 from database EMBL:
 human STS WI-11064.
 Score = 1523, P = 3.0e-64, identities = 327/343

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 50 bp to 2194 bp; peptide length: 715
 Category: similarity to known protein

```

1 MELADVGAAA SSQGVHDQVL PTPNASSRVI VHVLDLCFYA QVEMISNPGL
51 KDKPLGVQOK YLVVTCNYEA RKLGVKKLMN VRDAKEKCPQ LVLVNGEDLT
101 RYREMSYKVT ELLLEEFSPV ERLGFDFNFV DLTEMVEKRL QQLQSDLSA
151 VTVSGHYNNN QSINLLDVLH IRLLVGSQIA AEMREAMYNQ LGLTGCAGVA
201 SNKLLAKLVS GVFKPNQQT VLLPESCQHLI HSLNHIKEIP GIGYKTAKCL
251 EALGINSVRD LQTFSPKILE KELGISVAQR IQKLSFGEDN SPVILSGPPQ
301 SFSEEDSFKK CTSEVEAKNK IEELLASLLN RVCQDGRKPH TVRLIIRRY
351 SEKHYGRESR QCPIPSHVIQ KLGTYNDYDM TPMVDILMKL FRNMVNVKMP
401 FHLTLLSVCF CNLKALNTAK KGLIDYLYMP SLSTTSRSGK HSFKMKDTHM
451 EDFPKDKETN RDLFPGSGRI STRTRESPLD TTNFSKEKDI NEFPPLCSLPE
501 GVDQEVSKQL PVDIQEEILS GKSREKFQK GSVSCPLHAS RGVLSFFSKK
551 QMQDIPINPR DHLSSSKQVS SVSPCEPGTS GFNSSSSSYM SSQKDYSYYL
601 DNRKDKERIS QGKPEPQGFH FTNSNPVSA FHSEPNLQSE QLFERNHTTD
651 SHKQTVATDS HEGLTENREP DSVDEKITFP SDIDPQVFYE LPEAVQKELL
701 AEWKRTGSDF HIGHK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72b18, frame 2

PIR:H64747 DNA-damage-inducible protein dinP - Escherichia coli, N =
 2, Score = 212, P = 4.2e-27

PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis,
 N = 2, Score = 230, P = 5.2e-26

>PIR:H69963 DNA-damage repair protein homolog yqjH - Bacillus subtilis
 Length = 414

HSPs:

Score = 230 (34.5 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
 Identities = 47/112 (41%), Positives = 73/112 (65%)

Query: 27 SRVIVHVDLCFYAQVEMISNPGLKDKPLGV-----QQKYLVTTCNYEARKLGVKKLMNV 81
 SR+I H+D++ FYA VEM +P' L+ KP+ V ++K +VVTC+YEAR GVK M V
 Sbjct: 5 SRIIFHIDMNSFYASVEMAYDPALRGKPVAVAGNVKERRGIVVTCSEARARGVKTMPV 64

Query: 82 RDAKEKCPQLVLVNGEDLTRYREMSYKVTLEEFSPVVERLGFDFNFVDLTE 134
 AK CP+L+++ + RYR S + +L E++ +VE + DE ++D+T+
 Sbjct: 65 WQAKRHCPQLVLP-PNFDYRNSSRAMFTILREYTDLVEPVSIDEGYMDMTD 116

Score = 137 (20.6 bits), Expect = 5.2e-26, Sum P(2) = 5.2e-26
 Identities = 43/148 (29%), Positives = 75/148 (50%)

Query: 178 QIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQTLLPESCQHLIHSNLHIK 237
 + A E++ + +L L G+A NK LAK+ S + KP T+L ++ L +
 Sbjct: 125 ETAKETQSRQLKELLPSISIGIAPNKFALAKMASDMKKPLGITILRRQVPDILWPLP-VG 183

Query: 238 EIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSG 297
 E+ G+G KTA+ L+ LGI+++ +L L++ LGI+ R++ + G ++PV
 Sbjct: 184 EMHGVGKKTAEKLGKLGITIGELAAADEHSLKRLGGIN-GPRLKNKANGIHAPV---- 238

Query: 298 PPQSSEEDSFKKCTSEVEAKNKIEELL 325
 P+ E S ++ + EELL

Sbjct: 239 DPERIYEFKSVGNSSTLSHDSSDEEELL 266

Pedant information for DKFZphfbr2_72b18, frame 2

Report for DKFZphfbr2_72b18.2

[LENGTH] 715
 [MW] 80300.63
 [pI] 6.37
 [HOMOL] TREMBL:SPBC16A3_11 gene: "SPBC16A3.11"; product: "hypothetical protein";
 S.pombe chromosome II cosmid c16A3. 5e-30
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
 repair) [S. cerevisiae, YDR419w] 2e-15
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.
 genitalium, MG360] 3e-13
 [PIRKW] SOS mutagenesis 2e-11
 [PIRKW] DNA repair 2e-11
 [PIRKW] induced mutagenesis 2e-11
 [SUPFAM] umuC protein 3e-29
 [PROSITE] MYRISTYL 6
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 2
 [PROSITE] CK2_PHOSPHO_SITE 15
 [PROSITE] PROKAR_LIPOPROTEIN 1
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 21
 [PROSITE] ASN_GLYCOSYLATION 5
 [KW] Alpha Beta
 [KW] LOW_COMPLEXITY 4.20 %

SEQ MELADVGAASSQGVHDQVLPNASSRVIVHVDLDCFYAQVEMISNPELKDKPLGVQKQ
 SEG
 PRD ccc

SEQ YLVVTCNYEARLGLGVKKLMNVDRDAKEKCPQLVLVNGEDLTRYREMSYKVTELEEFSPVV
 SEG
 PRD ccc

SEQ ERLGFDENFVDLTEMVEKRLQQLQSDLSAVTVSGHVYNNQSNLLDVLHIRLLVGSQIA
 SEG
 PRD ecc

SEQ AEMREAMYNQLGLTGCAVASNKKLAKLVSGVFKPNQQTVLLPESQHLIHSNLHIKEIP
 SEG
 PRD hhhhhhhhhhhcc

SEQ GIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQ
 SEG
 PRD ccc

SEQ SFSEEDSFKKCTSEVEAKNKEELLASLLNRVCQDGRKPHTVRLIIRYSSEKHYGRESR
 SEG
 PRD ccc

SEQ QCPIPSHVIQKLGTYNDVMTMPVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAK
 SEG
 PRD ccc

SEQ KGLIDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLD
 SEG
 PRD hhhheeecc

SEQ TTNFSKEKDINEFPLCSLPEGVDQEVSKQLPVDIQEILSGKSREKFQKGKSVSCPLHAS
 SEG
 PRD ccc

SEQ RGVLSFFSKQMQDIPINPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSSQKDYSYL
 SEG
 PRD hcc

SEQ DNRLKDERISQGPKEPQGFHFTNSNPAVSAFHSFPNLQSEQLFSRNHTTDSHKQTVATDS
 SEG
 PRD hhhhhhhhhhhcc

SEQ HEGLTENREPDSVDEKITFPSDIDQVFYELPEAVQKELLAEWKRTGSDFHIGHK
 SEG
 PRD ccc

Prosites for DKFZphfbr2_72b18.2

PS00001	24->28	ASN_GLYCOSYLATION	PDOC00001
PS00001	160->164	ASN_GLYCOSYLATION	PDOC00001
PS00001	483->487	ASN_GLYCOSYLATION	PDOC00001
PS00001	583->587	ASN_GLYCOSYLATION	PDOC00001
PS00001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS00004	309->313	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	347->351	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	106->109	PKC_PHOSPHO_SITE	PDOC00005
PS00005	201->204	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	265->268	PKC_PHOSPHO_SITE	PDOC00005
PS00005	307->310	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	418->421	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	438->441	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	459->462	PKC_PHOSPHO_SITE	PDOC00005
PS00005	466->469	PKC_PHOSPHO_SITE	PDOC00005
PS00005	471->474	PKC_PHOSPHO_SITE	PDOC00005
PS00005	520->523	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	565->568	PKC_PHOSPHO_SITE	PDOC00005
PS00005	592->595	PKC_PHOSPHO_SITE	PDOC00005
PS00005	651->654	PKC_PHOSPHO_SITE	PDOC00005
PS00006	46->50	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	285->289	CK2_PHOSPHO_SITE	PDOC00006
PS00006	301->305	CK2_PHOSPHO_SITE	PDOC00006
PS00006	303->307	CK2_PHOSPHO_SITE	PDOC00006
PS00006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS00006	448->452	CK2_PHOSPHO_SITE	PDOC00006
PS00006	459->463	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	497->501	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	592->596	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	681->685	CK2_PHOSPHO_SITE	PDOC00006
PS00006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS00007	101->108	TYR_PHOSPHO_SITE	PDOC00007
PS00007	348->356	TYR_PHOSPHO_SITE	PDOC00007
PS00008	7->13	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	192->198	MYRISTYL	PDOC00008
PS00008	198->204	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	663->669	MYRISTYL	PDOC00008
PS00009	335->339	AMIDATION	PDOC00009
PS00013	186->197	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphfbr2_72b18.2)

DKFZphfbr2_72d13

group: brain derived

DKFZphfbr2_72d13 encodes a novel 165 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

seems to be testis specific 9 of 10 EST hits are from testis librarys

Sequenced by LMU

Locus: unknown

Insert length: 723 bp

Poly A stretch at pos. 704, no polyadenylation signal found

```
1 AGGGGGGGTA TGGGGGAGGG GGAGACTCTG CAGGAGCCTA ATTCCCCACT
51 CTGAGCTCAC CCTTCTGTCT GCCCGGGCCC TACCCCTTCC CCTACTCTCA
101 CCCTTATAAT CCTTTTCAGC ACTAGGTCTT CCCGTCACCT CCACCTCTCT
151 CCATGACCCG GCTCTGCTTA CCCAGACCCG AAGCACGTGA GGATCCGATC
201 CCAGTTCCCT CAAGGGGCCT GGGTGCTGGG GAGGGGTCAG GTAGTCCAGT
251 GCGTCCACCT GTATCCACCT GGGGCCCTAG CTGGGCCCAG CTCCTGGACA
301 GTGTCTCTATG GCTGGGGGCA CTAGGACTGA CAATCCAGGC AGTCTTTTCC
351 ACCACTGGCC CAGCCCTGCT GCTGCTTCTG GTCAGCTTCC TCACCTTTGA
401 CCTGCTCCAT AGGCCCGCAG GTCACACTCT GCCACAGCGC AAACCTTCTCA
451 CCAGGGGCCA GAGTCAGGGG GCCGCTGAAG GTCCTGGACA GCAGGAGGCT
501 CTACTCTGTC AAATGGGTAC AGTCTCAGGA CAACTTAGCC TCCAGGACGC
551 ACTGCTGCTG CTGCTCATGG GGCTGGGCCC GCTCCTGAGA GCCTGTGGCA
601 TGCCCTTGAC CTGCTTGCC CTGGCTTTCT GCCTCCATCC TTGGGCTGTA
651 GAGCCCTTCC CCACAACTCA GTGCTCTTCA AATATACAAT GACCACCTT
701 CTTCAAAAA AAAAAAAAAA AAC
```

BLAST Results

Entry HS860F19 from database EMBLNEW:
Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 860F19
Score = 2059, P = 1.1e-85, identities = 423/434
2 exons

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 153 bp to 647 bp; peptide length: 165
Category: putative protein
Classification: no clue

```
1 MTRLCLPRPE AREDPVPVPP RGLGAGEGSG SPVRPPVSTW GPSWAQLLDS
51 VLWLGLGLT IQAVFSTTGP ALLLLVLSFL TFDLLHRPAG HTLPQRKLLT
101 RGQSOGAGEG PGQOEALLQ MGTVSGQLSL QDALLLLMG LGPLLACGM
151 PLTLGLAFC LHPWA
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72d13, frame 3

No Alert BLASTP hits found

Report for DKFZphfbr2_72d13.3

```

SEQ      MTRLCCLRPEAREDDPIVPPRGLGAGEGSGSPVRPPVSTWGPSWAQLLDSVLWLGLALGLT
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhccccccc
MEM      .....

SEQ      IQAVFSTGTPALLLLLVSLFTDLLHRPAGHTLPQRKLLTRGQSQGAGEGPGQOEALLLQ
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      eeeeeccccchhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccchhhhhh
MEM      ..... MMMMMMMMMMMMMMMMMMMMM

SEQ      MGTVSGQLSLQDALLLLMLGLPLLRACGMPLTLLGLAFCLHPWA
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      hccccchhhhhhhhhhhhhhhccccchhhhhccccchhhhhhhhhccccccc
MEM      ..... MMMMMMMMMMMMMMMMMMMMM

```

(No Pfam data available for DKFZphfbr2_72d13.3)

DKFZphfbr2_72112

group: nucleic acid management

Summary DKFZphfbr2_72112 encodes a novel 344 amino acid protein with similarity to YDR126w and other *S. cerevisiae* proteins.

The novel protein contains a myc-type, helix-loop-helix dimerization domain signature. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers. Therefore, the protein could be a novel DNA-binding protein.

The new protein can application in modulating gene expression.

similarity to YDR126w ;
membrane regions: 2

similarity to YDR126w

complete cDNA complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 1270 bp
Poly A stretch at pos. 1251, no polyadenylation signal found

```
1 GGGGGCGCCC GGGAGGCGCC GGAGCCCAGC GGCTGGCGCC AGATCCAGGC
51 TCCTGGAAGA ACCATGTCCG GCAGCTACTG GTCATGCCAG GCACACACTG
101 CTGCCCCAAGA GGAGCTGCTG TTTGAATTAT CTGTGAATGT TGGGAAGAGG
151 AATGCCAGAG CTGCCGGCTG AAAATTACCC AACCAAGAGA AATCTGCAGG
201 ATGGACTTTC TGGTCCTCTT CTTGTTCTAC CTGGCTTCGG TGCTGATGGG
251 TCTGTCTCTT ATCTGCGTCT GCTCGAAAAC CCATAGCTTG AAAGGCCTGG
301 CCAGGGGAGG AGCACAGATA TTTTCCTGTA TAATTCCAGA ATGTCTTCAG
351 AGAGCCGTGC ATGGATTGCT TCATTACCTT TTCCATACGA GAAACCACAC
401 CTTCAATGTC CTGCACCTGG TCTTGCAAGG GATGGTTTAT ACTGAGTACA
451 CCTGGGAAGT ATTTGGCTAC TGTCAAGAGC TGGAGTTGTC CTTCATTAC
501 CTTCTTCTGC CCTATCTGCT GCTAGGTGTA AACCTGTTTT TTTTCACCTT
551 GACTTGTGGA ACCAATCCTG GCATTATAAC AAAAGCAAAT GAATTATTAT
601 TCTTTCATGT TTATGAATTT GATGAAGTGA TGTTCCTCAA GAACGTGAGG
651 TGCTCTACTT GTGATTTAAG GAAACCAGCT CGATCCAAGC ACTGCAGTGT
701 GTGTAACCTG TGTGTGCACC GTTTCGACCA TCACTGTGTT TGGGTGAACA
751 ACTGCATCGG GGCCTGGAAC ATCAGGTACT TCCTCATCTA CGTCTTGACC
801 TTGACGGCCT CGGCTGCCAC CGTCGCCATT GTGAGCACC ATTTTCTGGT
851 CCACTTGGTG GTGATGTCAG ATTTATACCA GGAGACTTAC ATCGATGACC
901 TTGGACACCT CCATGTTATG GACACGGTCA TTCTTATTCA GTACCTGTTC
951 CTGACTTTTC CACGGATTGT CTTCACTGCT GGCTTTGTCG TGGTCTCTAG
1001 CTTCTCCTG GGTGGCTACC TGTGTCTGT CCTGTATCTG GCGGCCACCA
1051 ACCAGACTAC TAACGAGTGG TACAGAGGTG TCTGGGCCTG GTGCCAGCGT
1101 TGTCCTCTTG TGGCCTGGCC TCCGTGAGCA GAGCCCCAAG TCCACCGGAA
1151 CATTCACTCC CATGGGCTTC GGAGCAACCT TCAAGAGATC TTTCTACCTG
1201 CCTTTCATG TCATGAGAGG AAGAAACAAG AATGACAAGT GTATGACTGC
1251 CAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1232 bp; peptide length: 344
Category: similarity to unknown protein


```

1 MDFLVLFIFY LASVLMGLVL ICVCSKTHSL KGLARGGAQI FSCIIEPECLQ
51 RAVHGLLHYL FHTRNHTFIV LHLVLQGMVY TEYTWEVFGY CQELELSLHY
101 LLLPYLLLVG NLFFFTLTCG TNPGIITKAN ELLFLHVEYF DEVMFPPKNVR
151 CSTCDLRKPA RSKHCSVCNW CVHRFDHHCW WVNNCIGAWN IRYFLIYVLT
201 LTASAATVAI VSTTFLVHLV VMSDLYQETY IDDLGHLHVM DTVILIQYLF
251 LTFPRIVFML GFVVVLSFLL GGYLLSVLYL AATNQTTNEW YRGVWAWCQR
301 CPLVAWPPSA EPQVHRNIHS HGLRSNLQEI FLPAFPCHER KKQE

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_72112, frame 3

TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1., N = 2, Score = 247, P = 1.4e-22

TREMBL:CED2021_3 gene: "D2021.2"; Caenorhabditis elegans cosmid
D2021., N = 1, Score = 209, P = 9e-17

TREMBL:CEC43H6_2 gene: "C43H6.7"; Caenorhabditis elegans cosmid
C43H6., N = 1, Score = 206, P = 5.2e-15

PIR:S52691 probable membrane protein YDR126w - yeast (Saccharomyces
cerevisiae), N = 1, Score = 207, P = 8.4e-15

PIR:E71607 metal binding protein (DHHC domain) PFB0725c - malaria
parasite (Plasmodium falciparum), N = 1, Score = 182, P = 1.1e-13

>TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein";
S.pombe chromosome II cosmid c13G1.
Length = 356

HSPs:

Score = 247 (37.1 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 55/148 (37%), Positives = 85/148 (57%)

Query: 52 AVHGLLHYLFHTRNH--TFIVLHLVLQGM---VYTEYTWEVFGYCQELELSLHYLLLPY 105
A+ L +Y+ + N F+ L L+ G+ +Y + F + + L +LLPY
Sbjct: 64 AMRSLSNYVLYKNNPLVVFLYLALITIGIASFFIYGSSLTQKFSIIDWISV-LTSVLLPY 122

Query: 106 LLLGVNLFFFTLTCGTNPGIITKANELLFLHVEYFD-EVMFPKNVRCSTCDLRKPARSKH 164
++L+ + +NPG I N + +D ++ FP +CSTC KPARSKH
Sbjct: 123 ----ISLY---IAAKSNPGKIDLNWNEASRRFPYDYKIFFPN--KCSTCKFEKPARSKH 173

Query: 165 CSVCNWCVHRFDHHCWVNNCIGAWNIRYFLIYVL 199
C +CN CV +FDHHC+W+NNC+G N RYF +++L
Sbjct: 174 CRLCNICVEKFDHHCWVNNCIGAWNIRYFLIYVL 208

Score = 43 (6.5 bits), Expect = 1.4e-22, Sum P(2) = 1.4e-22
Identities = 10/35 (28%), Positives = 17/35 (48%)

Query: 257 VFMLGFVV-VLSFLLGGYLLSVLYLAATNQTTNEW 290
VF++ + VL L GY ++Y T + +W
Sbjct: 254 VFLISLICSVLVLCLLGYEFLVYAGYTTNESEKW 288

Pedant information for DKFZphfbr2_72112, frame 3

Report for DKFZphfbr2_72112.3

```

[LENGTH]      344
[MW]           39677.23
[pI]           7.26
[HOMOL]        TREMBL:SPBC13G1_7 gene: "SPBC13G1.07"; product: "hypothetical protein"; S.pombe
chromosome II cosmid c13G1. 3e-17
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YDR126w] 1e-16
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YDR264c] 8e-05
[FUNCAT]       10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]
8e-05
[PIRKW]        transmembrane protein 4e-15
[SUPFAM]       ankyrin repeat homology 1e-10
[SUPFAM]       unassigned ankyrin repeat proteins 1e-10
[PROSITE]      MYRISTYL 4
[PROSITE]      CK2_PHOSPHO_SITE 3

```

```
{PROSITE}      PKC_PHOSPHO_SITE          1
{PROSITE}      ASN_GLYCOSYLATION        2
[KW]            SIGNAL_PEPTIDE          30
[KW]            TRANSMEMBRANE            2
[KW]            LOW_COMPLEXITY           16.57 %
```

[illegible]

Prosites for DKFZphfbr2_72112.3

PS000001	65->69	ASN_GLYCOSYLATION	PDOC000001
PS000001	284->288	ASN_GLYCOSYLATION	PDOC000001
PS000005	29->32	PKC_PHOSPHO_SITE	PDOC000005
PS000006	152->156	CK2_PHOSPHO_SITE	PDOC000006
PS000006	229->233	CK2_PHOSPHO_SITE	PDOC000006
PS000006	286->290	CK2_PHOSPHO_SITE	PDOC000006
PS000008	32->38	MYRISTYL	PDOC000008
PS000008	77->83	MYRISTYL	PDOC000008
PS000008	120->126	MYRISTYL	PDOC000008
PS000008	322->328	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2_72112.3)

DKFZphfbr2_72m16

group: unknown

DKFZphfbr2_72m16 encodes a novel 287 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 1462 bp

Poly A stretch at pos. 1441, polyadenylation signal at pos. 1421

```
1 GGGGAGGACC GGAGGACCGA GGACAGAAAG ATTGGTGGAC AGGAGCAGCG
51 GCCGGTGGGG AGGGCGCTCG GCGGCGGCCT GCGGCCATGG CCACCGTGAT
101 GGCAGCGACG GCGGCGGAGC GGGCGGTGCT GGAGGAGGAG TTCCGCTGGC
151 TGCTGCACGA CGAGGTGCAC GCTGTGTTGA AGCAGCTGCA GGACATCCTC
201 AAGGAGGCCT CTCTGCGCTT CACTCTGCCG GGCCTCCGCCA CTGAGGGGCC
251 CGCCAAGCAA GAGAACTTCA TCCTAGGCAG CTGTGGCACA GACCAGGTGA
301 AGGGTGTGCT GACTCTGCAG GGGGATGCCC TCAGCCAGGC GGATGTGAAC
351 CTGAAGATGC CCCGGAACAA CCAGCTGCTG CACTTCGCCT TCCGGGAGGA
401 CAAGCAGTGG AAGCTGCAGC AGATCCAGGA TGCCAGAAAC CATGTGAGCC
451 AAGCCATTTA CTTGCTTACC AGCCGGGACC AGAGCTACCA GTTCAAGACG
501 GCGCTGAGG TCCTCAAGCT GATGGACGCA GTGATGCTGC AGCTGACCAG
551 AGCCCCGAAAC CGGCTACCA CCCCCGCCAC CCTCACCCCTC CCCGAGATCG
601 CCGCCAGCGG CCTCACGCGG ATGTTGCCCC CTGCCCTGCC GTCCGACCTG
651 CTGGTCAACG TCTACATCAA CCTCAACAAG CTCTGCCTCA CGGTGTACCA
701 GCTGCATGCC CTGCAGCCCA ACTCCACCAA GAACTTCCGC CCAGCTGGGG
751 GCGCGGTGCT GCATAGCCCT GGGGCCATGT TCAGTGGGG CTCTCAGCGC
801 CTGGAGGTGA GCCACGTGCA CAAAGTGGAG TCGGTGATCC CCTGGCTCAA
851 CGACGCCCTG GTCTACTTCA CCGTCTCCCT GCAGCTCTGC CAGCAGCTTA
901 AGGACAAGAT CTCCGTGTTT TCCAGTACT GGAGCTACAG ACCCTTCTGA
951 TCACAGCACC CAGGAGCTTG TCTCCAGGAA GCGGGCCCCG TCCCTACTC
1001 ATACCACCA CAGAGCACA GCCAGTGCCA ACGCCAGGCT GCTATTTATC
1051 TCCCTATCCC ACCCCCTACC CCACCTAACA CATTTGCACT GCCGGGAATG
1101 GACACTGGAA GTGCCAGGAG GAAGGAAGGC TGGTTTGGTG GGGTAGTGGG
1151 GAGGTGAGG AGGCGGGGCC AAGGGTGTC CACATTCCCA ACACCGCCCT
1201 CTGATCACCA TGGGAATCTT TGGACTCAGG ACAGGGCCAG GCGCAGGGCT
1251 CTCCCTCCTC TCCCTTCGC TGTCCCTTCC CCCTGGAGGG CATGGTGTG
1301 GGGGGTGGCA CTGAGCTATG AGTCCCAGGG ATGGTGAGGA ACGCCACAGA
1351 CAGAGCCACC CTAGGAGTGA GTATAGTGCT GGTGACTGTG TTTCATAGCC
1401 CAGTCCAGG GCTGTCTAAG AAATAAAGAT CATCAGACTC CAAAAAATAA
1451 AAAAAAATAA AC
```

BLAST Results

Entry HS604351 from database EMBL:

human STS WI-18474.

Score = 1178, P = 1.5e-48, identities = 250/268

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 87 bp to 947 bp; peptide length: 287
Category: similarity to unknown protein

1	MATVMAATAA	ERAVLEEERF	WLLHDEVHAV	LKQLQDILKE	ASLRFTLPGS
51	GTEGPAKQEN	FILSCGCTDQ	VKGVLLTQSD	ALSQADVNLK	MPRNQLLHAF
101	AFREDKQWKL	QQIQDARNHV	SQAIIYLTGR	DQSYQFKFTA	EVLNKLDMVYM
151	LQLTRARNRL	TTPATLTLPK	IAASGLTRMF	APALPSDLIV	NVYINLNKLC
201	LTYYQLDALQ	PNSTKNFRPA	GGAVLHSPGS	MEFWGSQRLE	VSHVHKVECV
251	IPWLDLHALY	FVTSLQCCQ	LKDKISVFSS	YWSYRFF	

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2 72ml6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfbr2_72m16, frame 3

Report for DKFZphfbr2 72m16.3

```
[LENGTH]      287
[MW]           32254.40
[pI]           8.30
[HOMOL]       TREMBL:AF025459 2 gene: "H14A12.3"; Caenorhabditis elegans cosmid H14A12. 3e-14
```

[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	6
[PROSITE]	PKC_PHOSPHO_SITE	5
[PROSITE]	ASN_GLYCOSYLATION	1
[KW]	Alpha_Beta	
[KW]	LOW_COMPLEXITY	6.27 %

SEQ MATVMAATAAERAVLEEFRWLLHDEVHAVLKQLQDILKEASLRFTLPGSGTEGPAKQEN
SEG xxxxxxxxxxxxxxxxxxxx.....
PRD cccchhhcccccccccccchhhh

```
SEQ      FILGSCGTDQVKGVLTLQGDAISQADVNLMKPRNNQLLHFAFREDKQWKLQQIQDARNHV
SEG      .....
PRD      hhhccccccceeeeeeecccchhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
```

SEQ SQAIYLLTSRDQSYQFKTGAEVLKLDAMVQLTRARNRLTPATLTLPEIAASGLTRMF
SEG
PRD hhhhhhhhccccceecchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ APALPSDLLVNVYINLNKLCLTVYQLHALQPNSTKNFRPAGGAVLHSPGAMFEWGSQRLE
SEG
PRD cccccccceeeehhhhhhhhhhhheeeccccccccccccccccccccccccccccccce

```

SEQ      VSHVHKVECVI PWLNDALVYFTVSLQLCQQLKDKISVFSSYWSYRPF
SEG      .....
PRD      eeeeeeeeeeeccceeeeeehhhhhhhhhhhhhheeeeeeeccc

```

Prosites for DKFZphfbr2 72m16.3

PS000001	212->216	ASN_GLYCOSYLATION	PDOC000001
PS000005	42->45	PKC_PHOSPHO_SITE	PDOC000005
PS000005	128->131	PKC_PHOSPHO_SITE	PDOC000005
PS000005	213->216	PKC_PHOSPHO_SITE	PDOC000005
PS000005	236->239	PKC_PHOSPHO_SITE	PDOC000005
PS000005	283->286	PKC_PHOSPHO_SITE	PDOC000005
PS000006	8->12	CK2_PHOSPHO_SITE	PDOC000006
PS000006	50->54	CK2_PHOSPHO_SITE	PDOC000006
PS000006	83->87	CK2_PHOSPHO_SITE	PDOC000006
PS000006	128->132	CK2_PHOSPHO_SITE	PDOC000006
PS000006	138->142	CK2_PHOSPHO_SITE	PDOC000006
PS000006	167->171	CK2_PHOSPHO_SITE	PDOC000006
PS000008	64->70	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphfbr2 72m16.3)

DKFZphfbr2_72n12

group: brain derived

DKFZphfbr2_72n12 encodes a novel 117 amino acid protein with similarity to a protein with conserved sequence in bacteria and eukariota.

The novel protein is very similar to human MM46, human and rat ganglioside expression factor-2 (GEF2), C. elegans 14.8 kD protein C32D5.9 and Laccaria bicolor symbiosis-related protein LBU93506_1. The function of this highly conserved proteins is not known.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to rat GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="12"

Insert length: 1880 bp

Poly A stretch at pos. 1859, polyadenylation signal at pos. 1830

```
1 GGGGGCCGGT ATTTCTCCAT CTGGCTCTCC TCTACCTCCA GGCAGGCTCA
51 CCCGAGATCC CCGCCCCGAA CCCCCCTGTC ACACCTCGGC CAGCGCTGTT
101 GCGCCCGGAG CGGACGTTTC TGCAGCTATT CTGAGCACAC CTTGACGTCG
151 GCTGAGGGAG CGGGACAGGG TCAGCGGCGA AGGAGGCAGG CCCCgcgCGG
201 GGATCTCGGA AGCCCTGCGG TGCATCATGA AGTTCCAGTA CAAGGAGGAC
251 CATCCCTTTG AGTATCGGAA AAAGGAAGGA GAAAAGATCC GGAAGAAATA
301 TCCGGACAGG GTCCCCGTGA TTGTAGAGAA GGCTCCAAAA GCCAGGGTGC
351 CTGATCTGGA CAAGAGGAAG TACCTAGTGC CCTCTGACCT TACTGTTGGC
401 CAGTTCTACT TCTTAATCCG GAAGAGAATC CACCTGAGAC CTGAGGACGC
451 CTTATTCTTC TTGTCAACA ACACCATCCC TCCCACCACT GCTACCATGG
501 GCCAACGTGA TGAGGACAAT CATGAGGAAG ACTATTTTCT GTATGTGGCC
551 TACAGTGTAG AGAGTGCTTA TGGGAAATGA GTGGTTGGAA GCCCAGCAGA
601 TGGGAGCACC TGGACTTGGG GGTTAGGGGAG GGGTGTGTGT GCGCGACATG
651 GGGAAAGAGG GTGGCTCCCA CCGCAAGGAG ACAGAAGGTG AAGACATCTA
701 GAAACATTAC ACCACACACA CCGTCATCAC ATTTTCACAT GCTCAATTGA
751 TATTTTTTGC TGCTTCCTCG GCGGAGGAG AAAGCATGTC AGGACAGAGC
801 TGTTGGATTG GCTTTGATAG AGGAATGGGG ATGATGTAAG TTTACAGTAT
851 TCCTGGGGTT TAATTGTTGT GCAGTTTCAT AGATGGGTCA GGAGGTGGAC
901 AAGTTGGGGC CAGAGATGAT GGCAGTCCAG CAGCAACTCC CTGTGCTCCC
951 TTCTCTTTGG GCAGAGATTC TATTTTGTAC ATTTGCACAA GACAGGTAGG
1001 GAAAGGGGAC TTGTGGTAGT GGACCATACC TGGGGACCAA AAGAGACCCA
1051 CTGTAATTGA TGCATTGTGG CCCCTGATCT TCCCTGTCTC ACACCTCTTT
1101 TCTCCCATCC CGGTTGCAAT CTCACTCAGA CATCACAGTA CCACCCAGG
1151 GGTGGCAGTA GACAACAACC CAGAAATTTA GACAGGGATC TCTTACCTTT
1201 GGAAAATAGG GGTAGGCAT GAAGGTGGTT GTGATTAAGA AGATGGTTTT
1251 GTTATTAAAT AGCATTAAAC TGGAAATGAC AAGAGTGTG AGCATCCCTG
1301 TCTAACCTGC TCTTTCTCTT TGGTGCCCTT TATCTCACCC CTTCCTTGGA
1351 ATTTAATAAG TCTCAGGCAT TTCCAATTGT AGACTAAAAC CACTCTTAGC
1401 ATCTCTCTTA GTATTTTCCA TGATCAGGA AAGAGGTGTC TTATGTAGGG
1451 AGGGGGCAAG TATCAAGTAA GGTAATTATA TACTACTCTC ATTCAGGATT
1501 CTTGCTCCCA TGCTGCTGTC CCTTCAGGCT CACATGCACA GGAATGCTAC
1551 ATGATGGCCA GCTGCTTCCC TCCTTGGTTA TCATCCACTG CAGCTGCTAG
1601 TTAGAAAGGT TTGGAGGGAT GACTTTTAGT AAATCATGGG GATTTTATTG
1651 ATTTATTTTC ACTTTTGGGA TTTTGTGGG TGGGAGTGGG GAGCAGGAAT
1701 TGCACTCAGA CATGACATTT CAATTCATCT CTGCTAATGA AAAGGGTTCT
1751 TTCTCTTGGG GGAAATGTGT GTGTCAAGTC TGTCAGCTGC AAGTCTTGT
1801 ATAATGAAGT CAATGCCATC AGGCCAAGGA AATAAAATAA TTGTTACCT
1851 TAAAAATCGA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry HS418210 from database EMBL:

human STS SHGC-10496.

Score = 1916, P = 4.0e-80, identities = 394/400

Entry AC006514 from database EMBLNEW:

*** SEQUENCING IN PROGRESS *** Homo sapiens; HTGS phase 1, 68 unordered pieces.

Score = 610, P = 2.7e-16, identities = 128/134
4 exons

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 227 bp to 577 bp; peptide length: 117
Category: strong similarity to known protein

1 MKFYQKEDHP FEYRKKEGEK IRKKYPDRVP VIVEKAPKAR VPDLDKRKYL
51 VPSDLTVGQF YFLIRKRIHL RPEDALFFV NNTIPPTSAT MGQLYEDNHE
101 EDYFLYVAYS DESVYGK

BLASTP hits:

Entry YQD9_CAEEL from database SWISSPROT:
HYPOTHETICAL 14.8 KD PROTEIN C32D5.9 IN CHROMOSOME II.
Score = 496, P = 1.8e-47, identities = 91/116, positives = 105/116

Entry SYRP_LACBI from database SWISSPROT:
SYMBIOSIS-RELATED PROTEIN.
Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry LBU93506.1 from database TREMBL:
product: "symbiosis-related protein"; *Laccaria bicolor*
symbiosis-related protein mRNA, partial cds.
Score = 390, P = 3.1e-36, identities = 68/117, positives = 94/117

Entry GEF2_RAT from database SWISSPROT:
GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2).
Score = 373, P = 2.0e-34, identities = 71/116, positives = 88/116

Alert BLASTP hits for DKFZphfbr2 72n12, frame 2

TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete cds., N = 1. Score = 549, P = 4.7e-53

SWISSPROT:GEF2_HUMAN GANGLIOSIDE EXPRESSION FACTOR 2 (GEF-2)., N = 1,
Score = 373, P = 2.1e-34

```
>TREMBLNEW:AF044671_1 product: "MM46"; Homo sapiens MM46 mRNA, complete
  cds.
      Length = 117
```

HSPs :

Score = 549 (82.4 bits), Expect = 4.7e-53, P = 4.7e-53
Identities = 101/116 (87%), Positives = 110/116 (94%)

Query: 1 MKFYQYKEDHPFEYRKKEGEKIRKKYDRVPVIVEKAPKARVPLDKRKYLVPSDLTVGQF 60
MKF YKE+HPFE R+ EG EKIRKKYDRVPVIVEKAPKAR+ DLDD+KYLVP+SDLT+VGQF
Sbjct: 1 MKFVYKEEHPFEKRRSEGEKIRKKYDRVPVIVEKAPKARIGLDKKKYLVP+SDLT+VGQF 60

Query: 61 YFLIRKKRIHLRPEDALFFVFNNVNIPTPSATMGQLYEDNHEEDYFLYVAYSDESVDY 116
YFLIRKKRIHLR EDALFFVFNN IPTPSATMGQLY+++HEED+FLY+AYSDESVDY
Sbjct: 61 YFLIRKKRIHLRAEDALFFVFNNVNIPTPSATMGQLYQEHHEDDFLYVAYSDESVDY 116

Pedant information for DKFZphfbr2 72n12, frame 2

Report for DKFZphfbr2 72n12.2

```
[LENGTH]      117
[MW]           14044.07
[pI]           8.67
[HMOL]         TREMBL:AF044671.1 product: "MM46"; Homo sapiens MM46 mRNA, complete cds. 1e-56
```

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YBL078c] 4e-36
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YBL078c] 4e-36
[SUPFAM] hypothetical protein YBL078c 8e-35
[PROSITE] ASN_GLYCOSYLATION 1
[KW] Alpha_Beta

SEQ MKFQYKEDHPFEYRKKEGKIRKKYPDRVPVIVEKAPKARVPDLDRKKYLVPSDLTVGQF
PRD cccccccccchhhhhhhhhhhhhccccceeeccccccccccccccccccccccccchhh
SEQ YFLIRKRIHLRPEDALFFVNNITIPPTSATMGQLYEDNHEEDYFLYVAYSDESVMYK
PRD hhhhhhhhhccccceeeccccccccchhhhhhhhhccccceeecccccccccccc

Prosite for DKFZphfbr2_72n12.2

PS00001 81->85 ASN_GLYCOSYLATION PDOC00001

(No Pfam data available for DKFZphfbr2_72n12.2)

DKFZphfbr2_78c24

group: signal transduction

DKFZphfbr2_78c24 encodes a novel 563 amino acid protein with strong similarity to guanylate-binding proteins (GBPs).

GBPs were originally described as proteins that are strongly induced by interferons and are capable of binding to agarose-immobilized guanine nucleotides. hGBP1, the first of two members of this protein family in humans, represents a novel type of GTPase. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and a RGD cell attachment site. It seems to be a new member of the GBP-family and shows a splicing pattern not described previously.

The new protein can find application in modulating/blocking the response of cells to interferons.

strong similarity to guanine nucleotide-binding protein 1/2
but different "splice variant" aa 211-245 of GBP1/2 missing

Sequenced by MediGenomix

Locus: unknown

Insert length: 2952 bp

Poly A stretch at pos. 2927, polyadenylation signal at pos. 2914

```
1 CAGTTTCATT AGGCTCTGAA GCCATTACAA AGGTTGCTTA ACTTCTAATT
51 ATTTGATCAC TGAGGAAAAT CCAGAAAGCT ACACAACACT GAAGGGGTGA
101 AATAAAAAGTC CAGCGATCCA GCGAAAGAAA AGAGAAGTGA CAGAAACAAC
151 TTTACCTGGA CTGAAGATAA AAGCACAGAC AAGAGAACAA TGCCCTGGAC
201 ATGGCTCCAG AGATCCACAT GACAGGCCCA ATGTGCCTCA TTGAGAACAC
251 TAATGGGGAA CTGGTGGCGA ATCCAGAAGC TCTGAAAATC CTGTCTGCCA
301 TTACACAGCC TGTGGTGGTG GTGGCAATTG TGGGCCTCTA CCGCACAGGA
351 AAATCCTACC TGATGAACAA GCTAGCTGGG AAGAATAAGG GCTTCTCTCT
401 GGGCTCCACA GTGAAATCTC ACACCAAAGG AATCTGGATG TGGTGTGTGC
451 CTCACCCCAA AAAGCCAGAA CACACCTTAG TCCTGCTTGA CACTGAGGGC
501 CTGGGAGATG TAAAGAAGGG TGACAACCAG AATGACTCCT GGATCTTCAC
551 CCTGGCCGTC CTCCTGAGCA GCACTCTCGT GTACAATAGC ATGGGAACCA
601 TCAACCCAGA GGCTATGGAC CAACTGTACT ATGTGACAGA GCTGACACAT
651 CGAATCCGAT CAAAATCCTC ACCTGATGAG AATGAGAATG AGGATTCAGC
701 TGACTTTGTG AGCTTCTTCC CAGATTTTGT GTGGACACTG AGAGATTTCT
751 CCCTGGACTT GGAAGCAGAT GGACAACCCC TCACACCAGA TGAGTACCTG
801 GAGTATTTCC TGAAGCTAAC GCAAGGTAAC AGGAAGCTTG CCCAGCTTGA
851 GAAACTACAA GATGAAGAGC TGGACCCTGA ATTTGTGCAA CAAGTAGCAG
901 ACTTCTGTTC CTACATCTTT AGCAATTCCA AAACATAAAC TCTTTCAGGA
951 GGCATCAAGG TCAATGGGCC TTGTCTAGAG AGCCTAGTGC TGACCTATAT
1001 CAATGCTATC AGCAGAGGGG ATCTGCCCTG CATGGAGAAC GCAGTCTTGG
1051 CCTTGGCCCA GATAGAGAAC TCAGCCGCAG TGCAAAAGGC TATTGCCCAC
1101 TATGACCAGC AGATGGGCCA GAAGGTGCAG CTGCCCGCAG AAACCTTCCA
1151 GGAGCTGCTG GACCTGCACA GGGTTAGTGA GAGGGAGGCC ACTGAAGTCT
1201 ATATGAAGAA CTCTTTCAAG GATGTGGACC ATCTGTTTCA AAAGAAATTA
1251 GCGGCCACGC TAGACAAAAA GCGGGATGAC TTTTGTAAAC AGAATCAAGA
1301 AGCATCATCA GATCGTTGCT CAGCTTTACT TCAGGTCATT TTCAGTCTTC
1351 TAGAAGAAGA AGTGAAGGCG GGAATTTATT CGAAACCAGG GGGCTATTGT
1401 CTCTTTATTC AGAAGCTACA AGACCTGGAG AAAAAGTACT ATGAGGAACC
1451 AAGGAAGGGG ATACAGGCTG AAGAGATTCT GCAGACATAC TTGAAATCCA
1501 AGGAGTCTGT GACCGATGCA ATTCTACAGA CAGACCAGAT TCTCACAGAA
1551 AAGGAAAAGG AGATTGAAGT GGAATGTGTA AAAGCTGAAT CTGCACAGGC
1601 TTCAGCAAAA ATGGTGGAGG AAATGCAAAAT AAAGTATCAG CAGATGATGG
1651 AAGAGAAAGA GAAGAGTTAT CAAGAACATG TGAACAATTT GACTGAGAAG
1701 ATGGAGAGGG AGAGGGCCCA GTTGCTGGAA GAGCAAGAGA AGACCTCAC
1751 TAGTAAACTT CAGGAACAGG CCCGAGTACT AAAGGAGAGA TGCCAAGGTG
1801 AAAGTACCCA ACTTCAAAAT GAGATACAAA AGCTACAGAA GACCTGAAA
1851 AAAAAAACCA AGAGATATAT GTCGCATAAG CTAAAGATCT AAACAACAGA
1901 GCTTTTCTGT CATCCTAACC CAAGGCATAA CTGAAACAAT TTTAGAAATTT
1951 GGAACAAGTG TCACTATATT TGATAATAAT TAGATCTTGC ATCATAACAC
2001 TAAAAGTTTA CAAGAACATG CAGTTCAATG ATCAAAATCA TGTTTTCTCC
2051 TTAATAAGAT TGTAATTTGT GCAACAAAAG TGCATTTACC TCTGTACCAA
2101 CAGAGGAGGG ATCATGAGTT GCCACCACTC AGAAGTTTAT TCTTCCAGAC
2151 GACCACTGGA TACTGAGGAA AGTCTTAGGT AAAATCTTTG GGACATATTT
2201 GGGCACTGGT TTGGCCAAGT GTACAATAGG TCCCAATATC AGAACAACCC
2251 ATCCTAGCTT CCTAGGGAAG ACAGTGTAAC GTTCTCCATT ATATCAAGGC
2301 TACAAGGTCT ATGAGCAATA ATGTGATTTT TGGACATTGC CCATGGATAA
2351 TTTCTACTAG TGGATCTCAA GCTAAAGCAA ACCATCTTAT ACAGAGATCT
2401 AGAATCTTAT ATTTTCCATA GGAAGGTAAG GAAATCATTG GCAAGAGTAG
2451 GAATTGAATC ATAAACAAAT TGGCTAATGA AGAAATCTTT TCTTTCTTGT
2501 TCAATTCATC TAGATTATAA CCTTAATGTG ACACCTGAGA CCTTTAGACA
```



```

2551 GTTGACCTG AATTAAATAG TCACATGGTA ACAATTATGC ACTGTGTAAT
2601 TTTAGTAATG TATAACATGC AATGATGCAC TTTAACTGAA GATAGAGACT
2651 ATGTTAGAAA ATTGAACATA TTTAATTATT TGATTGTTT AATCCTAAAG
2701 CATAAGTTAG TCTTTTCCTG ATTCTTAAAG GTCATACTTG AAATCCTGCC
2751 AATTTTCCCC AAAGGGAATA TGGAAATTTT TTTGACTTTC TTTTGAGCAA
2801 TAAATAAATT GTCTTGCCAT TACTTAGTAT ATGTAGACTT CATCCCAATT
2851 GTCAAACATC CTAGGTAAGT GGTGACATT TCTTACAGCA ATTACAGATT
2901 ATTTTGAAC TAGAAATAAA CTAACTAGA AACAAAAAA AAAAAAAA
2951 AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 201 bp to 1889 bp; peptide length: 563
 Category: strong similarity to known protein
 Classification: Cell signaling/communication
 Prosite motifs: RGD (272-275)
 ATP_GTP_A (45-53)

```

1 MAPEIHMTGP MCLIENTNGE LVANPEALKI LSAITQPVVV VAIVGLYRTG
51 KSYLMNKLK KNGGSLGST VKSHTKGIWM WCVPHKKPE HTLVLLDTEG
101 LGDVKKGDQ NDSWIFTLAV LLSSTLVYNS MGTINQQAMD QLYYVTELTH
151 RIRSKSSPDE NENEDSADFV SFFPDFVWTL RDFSLEAD GQPLTPDEYL
201 EYSLKLTQGN RKLAQLEKLO DEELDPEFVQ QVADFCYIF SNSKTKTLTG
251 GIKVNGPCLE SLVLTYYINAI SRGDLPCMEN AVLALAQIEN SAAVQKAIH
301 YDQMGQKQV LPAETLQELL DLHRVSEREA TEVYMKNSEK DVDHLEFQKKL
351 AAQLDKKRD FCKQNEASS DRCSALLQVI FSPLEEVKA GIYSKPGGYC
401 LFIQKLQDL KKYEEPRKG IQAEELQTY LKSKESVTD ILQTDQILTE
451 KEKEIEVECV KAESAQASAK MVEEMQIKY QMEEKEKSY QEHVKQLTEK
501 MERERAQLE EQEKLTLTKL QEQARVLKER CQGESTQLQN EIQLKQTLK
551 KTKRMYMSHK LKI

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_78c24, frame 3

PIR:A41268 guanine nucleotide-binding protein 1 - human, N = 2, Score = 1306, P = 4.9e-238

PIR:A46459 macrophage-activation gene-1 protein mag-1 - mouse, N = 2, Score = 942, P = 8.9e-184

PIR:S70524 guanine nucleotide-binding protein 2 - human, N = 2, Score = 1131, P = 4.1e-210

TREMBL:AF077007_1 gene: "Gbp2"; product: "interferon-induced guanylate binding protein GBP-2"; Mus musculus interferon-induced guanylate binding protein GBP-2 (Gbp2) mRNA, complete cds., N = 2, Score = 904, P = 1.2e-179

>PIR:A41268 guanine nucleotide-binding protein 1 - human
 Length = 592

HSPs:

Score = 1306 (195.9 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
 Identities = 264/332 (79%), Positives = 288/332 (86%)

Query: 211 RKLAQLEKLQDEELDPEFVQVADFCYIFSNSKTKTLSSGGIKVNGPCLESVLTYINAI 270
 RKLAQLEKLQDEELDPEFVQVADFCYIFSNSKTKTLSSGGI+VNGP LESVLTY+NAI
 Sbjct: 245 RKLAQLEKLQDEELDPEFVQVADFCYIFSNSKTKTLSSGGIQVNGPRLESVLTYVNAI 304

Query: 271 SRGDLPCMENAVLALAQIENSAAVQKAIHYDQQMGQKVQLPAETLQELLDLHRVSEREA 330
S GDLPCMENAVLALAQIENSAAVQKAIHY+QMGQKVQLP E+LQELLDLHR SEREA
Sbjct: 305 SSGDLPCMENAVLALAQIENSAAVQKAIHYEQMGQKVQLPTESLQELLDLHRDSEREA 364

Query: 331 TEVYMKNSEFKDVLHFLQKLAQAQLDKKRDDFCQKQNEASSDRCSALLQVIFSPLEEEVKA 390
EV+++SFKDVLHFLQK+LAAQL+KKRDDFCQKQNEASSDRCS LLQVIFSPLEEEVKA
Sbjct: 365 IEVFIRSSFKDVLHFLQKELAAQLEKKRDDFCQKQNEASSDRCSGLLQVIFSPLEEEVKA 424

Query: 391 GIYSKPGGYCLFIQKLQDLKKYEEPRKGIQAEIILQTYLKSKESTDAILOTDQILTX 450
GIYSKPGGY LF+QKLQDL+KKYEEPRKGIQAEIILQTYLKSKESTDAILOTDQ LT
Sbjct: 425 GIYSKPGGYRLFVQKLQDLKKYEEPRKGIQAEIILQTYLKSKESTDAILOTDQTLTE 484

Query: 451 XXXXXXXXXXXXXSAQASAKMVEEMQIKYQQMMEEKESYQEHVKQLTEKMXXXXXXXX 510
SAQASAKM++EMQ K +QMMEE+KE+SYQEH+KQLTEKM
Sbjct: 485 KEKEIEVERVKAESAQASAKMLQEMQRKNEQMMEQKERSYQEHVKQLTEKMENDRVQLLK 544

Query: 511 XXXKTLTSLKLEQARVLKERCQGESTQLQNEI 542
+TL KLQEQ ++LKE Q ES ++NEI
Sbjct: 545 EQERTLALKLEQEQQLKEGFQKESRIMKNEI 576

Score = 1012 (151.8 bits), Expect = 4.9e-238, Sum P(2) = 4.9e-238
Identities = 194/211 (91%), Positives = 200/211 (94%)

Query: 1 MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLAV 60
MA EIHMTGPMCLIENTNG L+ANPEALKILSAITQP+VVVVAIVGLYRTGKSYLMNKLAV
Sbjct: 1 MASEIHMTGPMCLIENTNGRLMANPEALKILSAITQPMVVVAIVGLYRTGKSYLMNKLAV 60

Query: 61 KKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNDQNSWIFTLAV 120
K KGFSLGSTV+SHTKGIWMWCVPHPKK H LVLLDTEGLGDV+KGDNDQNSWIF LAV
Sbjct: 61 KKGFSLGSTVQSHTKGIWMWCVPHPKKPGHILVLLDTEGLGDVEKGDNDQNSWIFALAV 120

Query: 121 LLSSTLVYNSMGITINQAMQDQLYYVTELTTHIRSKSSPDENENE--DSADFVSFFPDFVW 178
LLSST VYNS+GTINQAMQDQLYYVTELTTHIRSKSSPDENENE DSADFVSFFPDFVW
Sbjct: 121 LLSSTFVYNSIGITINQAMQDQLYYVTELTTHIRSKSSPDENENEVEDSADFVSFFPDFVW 180

Query: 179 TLRDFSLEADGQPLTPDEYLEYSLKLTQG 209
TLRDFSLEADGQPLTPDEYL YSLKL +G
Sbjct: 181 TLRDFSLEADGQPLTPDEYLYSLKLKKG 211

Pedant information for DKFZphfbr2_78c24, frame 3

Report for DKFZphfbr2_78c24.3

[LENGTH] 563
[MW] 64127.72
[pI] 5.45
[HOMOL] PIR:A41268 guanine nucleotide-binding protein 1 - human 0.0
[SUPFAM] guanine nucleotide-binding protein 1 0.0
[PROSITE] ATP_GTP_A 1
[PROSITE] RGD 1
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 6.75 %
[KW] COILED_COIL 10.48 %

SEQ MAPEIHMTGPMCLIENTNGELVANPEALKILSAITQPVVVVAIVGLYRTGKSYLMNKLAV
SEG
PRD cccccccccccccccccchhhhhhhhhhhhhccceccccccccccccchhhhhhhh
COILS
MEMMMMMMMMMMMMMMMMM.....

SEQ KKGFSLGSTVKSHTKGIWMWCVPHPKKPEHTLVLLDTEGLGDVKKGDNDQNSWIFTLAV
SEG
PRD cccccccccccccccccceccccccccccccccccccccccccccccchhhhhhhh
COILS
MEM

SEQ LLSSTLVYNSMGITINQAMQDQLYYVTELTTHIRSKSSPDENENEEDSADFVSFFPDFVWTL
SEG
PRD hhhhhheccccchhhhhhhhhhhhhhhhhhhhhcccccceccccceeh
COILS
MEM

SEQ RDFSLEADGQPLTPDEYLEYSLKLTQGNRKLQLEKLQDEELDPEFVQQVADFCSYIF
SEG
PRD hhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhc
COILS

```

MEM .....
SEQ  SNSKTKTLSGGIKVNGPCLESVLTYINAIISRGDLPCMENAVLALAQIENSAAVQKAI AH
SEG  .....
PRD  cccceeeccccccccccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  YDQQMGQKVQLPAETLQELLDLHRVSEEREATEVYMKN SFKDVLH FQKKLAAQLDKKRDD
SEG  .....
PRD  hhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  FCKQNEASSDRCSALLQVIFSPLEEEVKAGIYSKPGGYCLFIQKLQDLEKKYYEPRKG
SEG  .....
PRD  hhhhhhhchhhhhhhhhhhhhhhhhhhhhhhccccccccceehhhhhhhhhhhhhcccccc
COILS .....
MEM  .....

SEQ  IQAEEILQTYLKSKE SVTDAILQTDQILTEKEKEIEVECVKAESAQASAKMVEEMQIKYQ
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....
MEM  .....

SEQ  QMEEKEKSYQEHVKQLTEKMERERAQLLEEQEKLTLSKLQEQARVLKERCQGESTQLQN
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
MEM  .....

SEQ  EIQLQLKTLKKKTKRYMSHKLKI
SEG  ..xxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCC.....
MEM  .....

```

Prosites for DKF2phfbr2_78c24.3

PS00016	272->275	RGD	PDOC00016
PS00017	45->53	ATP_GTP_A	PDOC00017

(No Pfam data available for DKF2phfbr2_78c24.3)

DKFZphfbr2_78d13

group: brain derived

DKFZphfbr2_78d13 encodes a novel 259 amino acid protein with similarity to *C. elegans* putative protein from cosmid K08B12.

No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to *C.elegans* K08B12.3

Sequenced by MediGenomix

Locus: /map="338.4 cR from top of Chr18 linkage group"

Insert length: 2195 bp

Poly A stretch at pos. 2175, polyadenylation signal at pos. 2156

```
1 CGTCCGTCGG GCAGCAGCGG GGCTGTCTAT CCCGGCTGAG GACCCGCGGC
51 CAGTGCCTGGT GGCTGGCTTT GCCATTAGCG GGGGCCTTTC CTGAGGACGG
101 CGTACGGAGT GTGGGGAATG AAGGATGGCA GCATGCCGTG CATTAAAAGC
151 TGTTTTGGTA GATCTCAGTG GCACACTTCA CATTGAAGAT GCAGCTGTGC
201 CAGGCGCACA GGAAGCTCTT AAAAGGTTAC GTGGTGCTTC TGTAATCATT
251 AGGTTTGTGA CCAATACAAC CAAAGAGAGC AAGCAAGACC TGTTAGAAAG
301 GTTGAGAAAA TTGGAATTTG ATATCTCTGA AGATGAAATA TTCACATCTC
351 TGACTGCAGC CAGAAAGTTA CTAGAGCGGA AACAAAGTCAG ACCCATGCTG
401 CTAGTTGATG ATCGGGCACT ACCTGATTTT AAAGGAATAC AAACAAGTGA
451 TCCTAATGCT GTGGTCATGG GATTGGCACC AGAACATTTT CATTATCAAA
501 TTCTGAATCA AGCATTCCGG TTACTCCTGG ATGGAGCACC TCTGATAGCA
551 ATCCACAAAG CCAGGTATTA CAAGAGGAAA GATGGCTTAG CCCTGGGGCC
601 TGGACCATTT GTGACTGCTT TAGAGTATGC CACAGATACC AAAGCCACAG
651 TCGTGGGGAA ACCAGAGAAG ACGTTCTTTT TGAAGCATT GCGGGGCACT
701 GGCTGTGAAC CTGAGGAGGC TGTCTGATA GGAGATGATT GCAGGGATGA
751 TGTGGTGGGG GCTCAAGATG TCGGCATGCT GGGCATCTTA GTAAAGACTG
801 GGAATATCG AGCATCAGAT GAAGAAAAAA TTAATCCACC TCCTTACTTA
851 ACTTGTGAGA GTTCCCTCA TGTGTGGAC CACATTCTGC AGCACCTATT
901 GTGAAGCAAT GTGTGCATCT GAAGCAACTT GAAATGCAGC TTCTTATTGT
951 CTGGAATGAA TCCCTTACCA ACTCAGTGCC AGCATCGGTA GACACCAGTC
1001 AGTGCTGATC GCTTTTAAAC CCTCTTTTGT TGTGCATTAA TTAGAAAGAA
1051 AGGTATTGAA TTGCGGCTAG CCAGTAAGCC TTGCTAATCT CTTTATTTTT
1101 GTAACCTGAG ATGAGACCCA AAGAAAGGGA AAGCTGAGAT TTTGTGCCAT
1151 TCCTTTTAAA ATATTCATCA GGTAGGTGG GGCTGTGGGG GAAAAGCTAC
1201 TACAGGGAAG AGTGTCTCT GCTGTCTCT CACTGAAAAA CAGGGAGGGG
1251 GGATTTCAGA CTGTGAAGAA AGTTGAATGG TGGTTTTTAA ATTATAAAGT
1301 AATGTATTAA AAGGTGCATT AGGCTGTAGT TCTAATATTG AGTTCAACTG
1351 TGAATCCAT CAGATGTGCC AAATGGAGAA GACAGAAAGC AACAAAGTGA
1401 ATTGTTCTTT AGCCCAAGTG GTACAGTGAA TTTGCTTTAA CAGATGTTGA
1451 AAATAAATT TTCTACTGTA TTCCAGCAC GGGTGACTTC TTTTCTCTTT
1501 CATTAGCCAG AGATGACTAA TTAAATTTA GAACCAGATT TTAATTTAAA
1551 TTAATATTTT CATTAAATAC CTACTCATTG CAGATACCTA TTATACTGTG
1601 TAACAGTTGT TTTGGAAATT TTATGTAAAA TTAATACTAT CAGTATTTTA
1651 CAGATGTTTT AATTAGACAT TGTATTAAAC AGGAACAGTG CAGAACTAG
1701 AATCAAGCCT TATAATATCT TATAGACCAT GCATTTTGA AGTTAGTGTC
1751 CACTAGGGTC CTATTAACCTG TACATTGCA AGATTTTATT ATTTTGCCTT
1801 CTGACACTAT GGGAAAAATT TTTAGAAGC TATTGGGACA GATTCAAGCT
1851 TTTATGCACT TGGTTACTAC AGCTGTAAAA TGAATCTCG TCTTGTAGCA
1901 TGGATTATTC TTCTCATGTT AAACCCACCA AAATAAAGGG GACTAAATAG
1951 GTAATGATTT TCCTAGTGCA TTTGCATACT GTGATAATCC TGGGCCTTGC
2001 AATAGTTCTA CAGGGCTCTT GGGCATTGAA TTATTAGGAT GTAATTGTAC
2051 ATCATTGTAG TGTTCACCTT ATTGAAGCTC ACTCTGATGT TAATGAGCTT
2101 CGGGTTTTGA TGCTTGTTTA GAGATCAGCA GTCTTGGATG GGAGGGAACA
2151 AAGCTAAATA AATGTTAGTT TGGTGAAAAA AAAAAAAAAA AAAAA
```

BLAST Results

Entry HS599355 from database EMBL:

human STS WI-13484.

Score = 1262, P = 3.6e-52, identities = 274/289

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 125 bp to 901 bp; peptide length: 259
Category: similarity to unknown protein
Classification: no clue

```

1 MAACRALKAV LVDLSGTLHI EDAAVPGAQE ALKRLRGASV IIRFVTNTTK
51 ESKQDLLERL RKLEFDISE EIFTSLTAAR SLLERKQVRP MLLVDDRALP
101 DFKGIQTSDF NAVVMGLAPE HFHYQILNQA FRLLLDGAPL IAIHKARYYK
151 RKDGLALGPG PFVTALEYAT DTKATVVGKP EKTFFLEALR GTGCEPEEAV
201 MIGDDCRDDV GGAQDVGMLG ILVKTGKYRA SDEEKINPPP YLTCEFPFHA
251 VDHILQHLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78d13, frame 2

TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12., N = 1, Score = 609, P = 2.2e-59

TREMBL:CEC13C4_5 gene: "C13C4.4"; Caenorhabditis elegans cosmid C13C4, N = 1, Score = 408, P = 4.4e-38

>TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12.
Length = 257

HSPs:

Score = 609 (91.4 bits), Expect = 2.2e-59, P = 2.2e-59
Identities = 132/251 (52%), Positives = 172/251 (68%)

```

Query: 7 LKAVLVDSGLTHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERLRKLEFD 66
      + +VL+DLSGT+HIE+ A+PGAQ AL+ LR + + +FVTNTTKESK+ L +RL F
Sbjct: 4 ISSVLIDLSGTIHIEEFAIPGAQTALELLRQHAKV-KFVTNTTKESKRLHQRLLINCGFK 62

Query: 67 ISEDEIFTSLTAARSLLERKQVRPMLLVDDRALPDFKGIQTSDFNAVVMGLAPEHFHYQI 126
      + ++EIFTSLTAAR L+ + Q RP +VDDRA+ DF+GI T DPNVAV+GLAPE F+
Sbjct: 63 VEKEEIFTSLTAARDLIVKNQYRPFVIVDDRAMEDFEGISTDDPNVAVIGLAPEKFNDDT 122

Query: 127 LNQAIFRLLLDG-APLIAIHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKPEKTF 185
      L AFRL+ + A LIAI+K RY++ GL LGPG +V LEY+ +AT+VGKP K FF
Sbjct: 123 LTHAFRLIKEKKASLIAINKGRYHQTNAGLCLGPGCTYVAGLEYSAGVEATIVGKPNKLF 182

Query: 186 LEALRGTG--CEPEEAVMIGDDCRDDVGGAQDVGMLGILVKTGKYRASDEEKNPPPYLT 243
      AL+ + AVMIGDD DD GA +GM ILVKTGK+R DE K+
Sbjct: 183 ESALQSLNENVDFSSAVMIGDDVNDALGAIKGMRAILVKTGKFRDGDDELKVKN----V 238

Query: 244 CESFPHAVDHILQH 257
      SF AV+ I+++
Sbjct: 239 ANSFVDAVNMIEN 252

```

Pedant information for DKFZphfbr2_78d13, frame 2

Report for DKFZphfbr2_78d13.2

```

[LENGTH] 259
[MW] 28536.04
[pI] 5.84
[HOMOL] TREMBL:CEUK08B12_1 gene: "K08B12.3"; Caenorhabditis elegans cosmid K08B12. 3e-62
[FUNCAT] r general function prediction [M. jannaschii, MJ1437] 3e-05
[SUPFAM] nagD protein 4e-18
[KW] Alpha_Beta

```

```

SEQ  MAACRALKAVLVDLSGTLHIEDAAVPGAQEALKRLRGASVIIRFVTNTTKESKQDLLERL
PRD  cccccccccccccccccccccccccchhhhhhhhhccccccccccccchhhhhhhhh

SEQ  RKLEFDISEDEIFTSLTAARSLERKQVRPMLLVDDRALPDFKGIQTSDPNAVVMGLAPE
PRD  hhhccccccccceehhhhhhhhhhhccccccccceehhhhhhhccccccccccccceehhhhhhh

SEQ  HFHYQILNQAFRLLLDGAPLIAIHKARYYKRKDGALGPGPFVTALEYATDTKATVVGKP
PRD  chhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhccccccccccccceehhhhhhh

SEQ  EKTFFLEALRGTCPEEAVMIGDDCRDDVGGAQDVGMGLILVKTGKYRASDEEKINPPP
PRD  cchhhhhhhhhhhccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccc

SEQ  YLTCESFPHAVDHILQHLL
PRD  cccccchhhhhhhhhhhcccc

```

(No Prosite data available for DKF2phfbr2_78d13.2)

(No Pfam data available for DKF2phfbr2_78d13.2)

DKFZphfbr2_78k24

group: metabolism

DKFZphfbr2_78k24 encodes a novel 372 amino acid protein with similarity to *Mus musculus* ubiquitin specific protease UBP43.

The novel protein contains a Prosite ubiquitin carboxyl-terminal hydrolases family 2 signature 2. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins.

The new protein can find application in modulation of protein stability/degradation in cells.

Ubiquitin carboxyl-terminal hydrolases family 2 signature 2.

strong similarity to mouse ubiquitin specific protease UBP43

Sequenced by MediGenomix

Locus: unknown

Insert length: 1874 bp

Poly A stretch at pos. 1852, polyadenylation signal at pos. 1836

```
1 AGTCCCAGC TGGAACTCAG CAGCGGAGGC TGGACGCTTG CATGGCGCTT
51 GAGAGATTCC ATCGTGCCCTG GCTCACATAA GCGCTTCCTG GAAGTGAAGT
101 CGTGCTGTCC TGAACGCGGG CCAGGCAGCT GCGGCCTGGG GGTTTTGGAG
151 TGATCACGAA TGAGCAAGGC GTTTGGGCTC CTGAGGCAAA TCTGTCAATC
201 CATCCTGGCT GAGTCCTCGC AGTCCCCGGC AGATCTTGAA GAAAAGAAGG
251 AAGAAGACAG CAACATGAAG AGAGAGCAGC CCAGAGAGCG TCCCAGGGCC
301 TGGGACTACC CTCATGGCCT GGTGTTTGA CACAACATTG GACAGACCTG
351 CTGCCTTAAC TCCTTGATTC AGGTGTTCGT AATGAATGTG GACTTCACCA
401 GGATATTGAA GAGGATCACG GTGCCAGGGG GAGCTGACGA GCAGAGGAGA
451 AGCGTCCCTT TCCAGATGCT TCTGCTGCTG GAGAAGATGC AGGACAGCCG
501 GCAGAAAGCA GTGCGGCCCC TGGAGCTGGC CTAAGCTGCT CAGAGTGAAC
551 ACGTGCCCTT GTTTGTCCAA CATGATGCTG CCCAAGTGTG CCTCAAACTC
601 TGGAACTCTG TTAAGGACCA GATCACTGAT GTGCACTTGG TGGAGAGACT
651 GCAGGCCCTG TATACGATCC GGGTGAAGGA CTCCTTGATT TGCGTTGACT
701 GTGCCATGGA GAGTAGCAGA AACAGCAGCA TGCTCACCCCT CCCACTTTCT
751 CTTTTTGATG TGGACTCAAA GCCCCTGAAG ACACTGGAGG ACGCCCTGCA
801 CTGCTTCTTC CAGCCAGGGG AGTTATCAAG CAAAAGCAAG TGCTTCTGTG
851 AGAACTGTGG GAAGAAGACC CGTGGGAAAC AGGTCTTGAA GCTGACCCAT
901 TTGCCCCAGA CCTGACAAT CCACCTCATG CGATTCTCCA TCAGGAATTC
951 ACAGACGAGA AAGATCTGCC ACTCCCTGTA CTTCCTCCAG AGCTTGATT
1001 TCAGCCAGAT CCTTCCAATG AAGCGAGAGT CTTGTGATGC TGAGGAGCAG
1051 TCTGGAGGGC AGTATGAGCT TTTTGCTGTG ATTGCGCAGC TGGGAATGGC
1101 AGACTCCGGT CATTACTGTG TCTACATCCG GAATGCTGTG GATGGAAAAT
1151 GGTTCCTGCT CAATGACTCC AATATTGCTT TGGTGTCTG GGAAGACATC
1201 CAGTGTACCT ACGGAAATCC TAACTACCAC TGGCAGGAAA CTGCATATCT
1251 TCTGGTTTAC ATGAAGATGG AGTGCTAATG GAAATGCCCA AAACCTTCAG
1301 AGATTGACAC GCTGTCAATT TCCATTTCGG TTCTTGATC TACGGAGTCT
1351 TCTAAGAGAT TTTGCAATGA GGAGAAGCAT TGTTCCTAAA CTATATAACT
1401 GAGCCTTATT TATAATTAGG GATATTATCA AAATATGTAA CCATGAGGCC
1451 CCTCAGGTCC TGATCAGTCA GAATGGATGC TTTCACACAG AGACCCGGCC
1501 ATGTGGCTGC TCGGTCCTGG GTGCTCGCTG CTGTGCAAGA CATTAGCCCT
1551 TTAGTTATGA GCCTGTGGGA ACTTCAGGGG TTCCAGTGGG GGAGAGCAGT
1601 GGCAGTGGGA GGCATCTGGG GGCCAAAGGT CAGTGGCAGG GGGTATTTC
1651 GTATTATACA ACTGCTGTGA CCAGACTTGT ATACTGGCTG AATATCAGTG
1701 CTGTTGTGTA TTTTCACTT TGAGAACCAA CATTAAATCC ATATGAATCA
1751 AGTGTGTTGT AACTGCTATT CATTTATTCA GCAAAATTTT ATTGATCATC
1801 TCTTCTCCAT AAGATAGTGT GATAAACACA GTCATGAATA AAGTTATTTT
1851 CCACAAAAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry AC005500 from database EMBL:

, complete sequence.

Score = 859, P = 5.7e-143, identities = 175/179

8 exons matching Bp 317-1230

Medline entries

99182491:

A novel ubiquitin-specific protease, UBP43, cloned from leukemia fusion protein AML1-ETO-expressing mice, functions in hematopoietic cell differentiation.

Peptide information for frame 1

ORF from 160 bp to 1275 bp; peptide length: 372
 Category: strong similarity to known protein
 Classification: Protein management
 Prosite motifs: UCH_2_2 (302-320)

```

1 MSKAFGLLRQ ICQSILAESS QSPADLEEEK EEDSNMKREQ PRERPRAWDY
51 PHGLVGLHNI GQTCCNLNLI QVFVMNVDFTRILKRITVPR GADEQRRSVP
101 FQMLLLLEKM QDSRQKAVRP LELAYCLQKC NVPLFVQHDA AQLYLKLWNL
151 IKDQITDVHL VERLQALYTI RVKDSLICVD CAMESSRNSS MLTLPPLSLFD
201 VDSKPLKTLE DALHCFQFQPR ELSSKSKCFC ENCGKKTRGK QVLKLTHLPQ
251 TLTIHLMRFS IRNSQTRKIC HSLYFPQSLD FSQILPMKRE SCDAAEQSGG
301 QYELFAVIAH VGMADSGHYC VYIRNAVDGK WFCFNDNIC LVSWEDIQCT
351 YGNPNYHWQE TAYLLVYMKM EC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_78k24, frame 1

TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds., N = 1, Score = 1367, P = 1e-139

SWISSPROT:UBPE_DROME UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 64E (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 64E) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 64E) (DEUBIQUITINATING ENZYME 64E)., N = 2, Score = 248, P = 5.3e-33

>TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus ubiquitin specific protease UBP43 mRNA, complete cds.
 Length = 368

HSPs:

Score = 1367 (205.1 bits), Expect = 1.0e-139, P = 1.0e-139
 Identities = 262/369 (71%), Positives = 295/369 (79%)

```

Query:      1 MSKAFGLLRQICQSILAESSQSPADLEEEK EEDSNMKREQPRERPRAWDYPHGLVGLHNI 60
            M K FGLLR+ CQS++AE Q A LEE E   KR  R+  AWD PHGLVGLHNI
Sbjct:      1 MGKGFGLLRKPQSVVAEPQQYSA-LEE--ERTMKRKRVLSDLCSAWDSPHGLVGLHNI 57

Query:      61 GQTCCNLNLIQVFVMNVDFTRILKRITVPRGADEQRRSVPFQMLLLLEKMQDSRQKAVRP 120
            GQTCCNLNLI+QVF+MN+DF  ILKRITVPR A+E++RSVPFQ+LLLLLEKMQDSRQKA+ P
Sbjct:      58 GQTCCNLNLIQVFMMNMDFRMILKRITVPRSAEERKRVPFQLLLLLEKMQDSRQKALLP 117

Query:      121 LELAYCLQKCNVPLFVQHDAQAQLYLKLWNLIKDQITDVHLVERLQALYTIIRVKDSLICVD 180
            EL  CLQK NVPLFVQHDAQAQLYL +WNL KDQITD L ERLQ L+TI ++SLICV
Sbjct:      118 TELVQCLQKYNVPLFVQHDAQAQLYLTIWNLTKDQITDITDLTERLQGLFTIWTQESLICVG 177

Query:      181 CAMESSRNSSMLTLPPLSLFDVDSKPLKTLEDALHCFQFQPRELSSKSKCFCENCGKKTRGK 240
            C ESSR S +LTL L LFD D+KPLKTLEDAL CF QP+EL+S  C CE CG+KT K
Sbjct:      178 CTAESSRRSKLLTSLPLFDKDAKPLKTLEDALRCFVQPKELASSDMC-CETCGEKTTPWK 236

Query:      241 QVLKLTHLPQTLTIHLMRFSIRNSQTRKICHSLYFPQSLDIFSQILPMKRESQDAEQSGG 300
            QVLKLTHLPQTLTIHLMRFS RNS+T KICHS+ FPQSLDIFSQ+LP + + D +EQS
Sbjct:      237 QVLKLTHLPQTLTIHLMRFSARNRSTKICHSVNFPQSLDIFSQVLPTEEDLGDTKEQSEI 296

Query:      301 QYELFAVIAHVGMADSGHYCVYIRNAVDGKWFCFNDNICLVSWEDIQCTYGNPNYHWQE 360
            YELFAVIAHVGMAD GHYC YIRN VDGKWFCFNDNS++C V+W+D+QCTYGN Y W+E
Sbjct:      297 HYELFAVIAHVGMADFGHYCAYIRNPVDGKWFCFNDSHVCVWTWKDVQCTYGNHRYRWRE 356

Query:      361 TAYLLVYMK 369

```


Sbjct: 357 TAYLLVYTK 365

Pedant information for DKFZphfbr2 78k24, frame 1

Report for DKFZphfbr2 78k24.1

```
[LENGTH] 372
[MW] 43011.12
[pI] 8.05
[HOMOL] TREMBLNEW:AF069502_1 product: "ubiquitin specific protease UBP43"; Mus musculus
ubiquitin specific protease UBP43 mRNA, complete cds. 1e-151
[FUNCAT] 06.13 proteolysis [S. cerevisiae, YMR304w] 3e-19
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YJL197w] 3e-16
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YMR223w] 1e-15
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL186w] 6e-12
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR069c] 9e-11
[FUNCAT] 10.03.99 other osmosensing activities [S. cerevisiae, YDR069c] 9e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR069c] 9e-11
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR069c] 9e-11
[FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YDR069c] 9e-11
[BLOCKS] BL00582A Ribosomal protein L33 proteins
[BLOCKS] BL00972E
[BLOCKS] BL00972D
[BLOCKS] BL00972A
[EC] 2.4.2.29 Queuine tRNA-ribosyltransferase 1e-06
[PIRKW] pentosyltransferase 1e-06
[PIRKW] glycosyltransferase 1e-06
[PIRKW] tRNA modification 1e-06
[PIRKW] alternative splicing 7e-11
[PIRKW] hydrolase 7e-06
[SUPFAM] deubiquinating enzyme SSV7 2e-09
[PROSITE] UCH_2_2_1
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[PFAM] Ubiquitin carboxyl-terminal hydrolases family 2
[KW] Alpha Beta
```

SEQ	MSKAFGLLRQICQSILAESSQSPADLEEKKEEDSNMKREQPRERPAWDYPHGLVGLHNI
PRD	ccccceehhhhhhhhhccccccccchhhhhhhhhcccccccccccccccccccccccccccc
SEQ	GQTCCLNSLIQVFMVNDFTRI LKRITVPRGADEQRRSVFPQMLLLEKMQDSRQKAVRP
PRD	cceehhhhhhhhhhhccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhcccc
SEQ	LEHAYCLQKCNVLVQFHDAQAQLYKLWNLIKQDITDVLHVERLQALYIRVKDSLICVD
PRD	hhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhh
SEQ	CAMESSRNSMMLTLP LSLFDVDSKPLKLTLEDALHCFQPRELSSSKSCFCENCCKKTRGK
PRD	ccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccecccccccccc
SEQ	QVLKLTHLPQTLTIHLMRFSIRNSQTRKICHSLYFPQS LDFSQILPMKRESQDAEEQSGG
PRD	ccceeeccccchhhhhhhhhhhhhccccchhhhhcccccccccccccccccccccccccccccc
SEQ	QYELFAVIAHVGMADSGHYCVYIRNAVDGKWFCDNDSNICLVSWEDIQCTYGNPNYHWQE
PRD	eeeeeeeeeeeeccccccceeeeeeeccccccccceeeccccceeeccccccccccccccccchh
SEQ	TAYLLVYMKMEC
PRD	hhhhhhhhhhcccc

Prosite for DKF2phfbr2 78k24.1

PS00973 302->320 UCH 2 2 PDOC00750

Pfam for DKFZphfbr2 78k24.1

HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*GIqNlGNTCYMNSIIQCL*		
	G+ N+G TC +NS+IQ+		
Query	56	GLHNIGQTCCLSLSIQVF	73

```
HMM_NAME      Ubiquitin carboxyl-terminal hydrolases family 2
HMM            *YdLYgVICHYGntldyGHYWaYVKNenhHRWkWYYFDDEtV*
               Y+L++VI H G  D+GHY +Y++N  ++KW++F+D+++
Query          302 YELFAVIAHVG-MADSGHYCVYIRNAV--DGKWFCFNDSTI 339
```

DKFZphfbr2_78n23

group: brain derived

DKFZphfbr2_78n23 encodes a novel 329 amino acid protein with similarity to A.thaliana F26P21.80 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F26P21.80

Sequenced by MediGenomix

Locus: /map="89.1 cR from top of Chr19 linkage group"

Insert length: 1447 bp

Poly A stretch at pos. 1374, polyadenylation signal at pos. 1353

```
1 TACAACTTCC GGCTGTAAAG ATGGCGGCTT CCTAGTGAGT CGGCGGCTGA
51 CTTAGAAGGA GGTTCAGGCT ACGGTGAGCC GAAGCCACAC AGGAGCCATG
101 GAAAGTGGCAG AGCCCAGCAG CCCCACTGAA GAGGAGGAGG AGGAAGAGGA
151 GCACTCGGCA GAGCCTCGGC CCCGCACTCG CTCCAATCCT GAAGGGGCTG
201 AGGACCGGGC AGTAGGGGCA CAGGCCAGCG TGGGCAGCCG CAGCGAGGGT
251 GAGGGTGAGG CCGCCAGTGC TGATGATGGG AGCCTCAACA CTTCAGGAGC
301 CGGCCCTAAG TCCTGGCAGG TGCCCCGCC AGCCCCTGAG GTCCAAATTC
351 GGACACCAAG GGTCAACTGT CCAGAGAAAG TGATTATCTG CCTGGACCTG
401 TCAGAGGAAA TGTCACTGCC AAAGCTGGAG TCGTTCAACG GCTCCAAAAC
451 CAACGCCCTC AATGTCTCTC AGAAGATGAT TGAGATGTTT GTGCGGACAA
501 AACACAAGAT CGACAAAAGC CACGAGTTTG CACTGGTGGT GGTGAACGAT
551 GACACGGCCT GGCTGTCTGG CCTGACCTCC GACCCCGCGG AGCTCTGTAG
601 CTGCCTCTAT GATCTGGAGA CGGCCTCCTG TTCCACCTTC AATCTGGAAG
651 GACTTTTTCAG CCTCATCCAG CAGAAAACCT AGCTTCCGGT CACAGAGAAC
701 GTGCAGACGA TTCCCCGCC ATATGTGGTC CGCACCATCC TTGTCTACAG
751 CCGTCCACCT TGCCAGCCCC AGTTCTCCTT GACGGAGCCC ATGAAGAAAA
801 TGTTCAGTG CCCATATTTC TTCTTTGACG TTGTTTACAT CCACAATGGC
851 ACTGAGGAGA AGGAGGAGGA GATGAGTTGG AAGGATATGT TTGCCTTCAT
901 GGGCAGCCTG GATACCAAGG GTACCACTA CAAGTATGAG GTGGCACTGG
951 CTGGGCCAGC CCTGGAGTTG CACAACCTGCA TGGCGAAACT GTTGGCCAC
1001 CCCCTGCAGC GGCCTTGCCA GAGCCATGCT TCCTACAGCC TGCTGGAGGA
1051 GGAGGATGAA GCCATTGAGG TTGAGGCCAC TGCTGAACC ATCCCTGTAC
1101 ATCTGCACCT TCTGTGCAA GGAAGTCCTT GGCCTAAAGC CTTGGTTCTC
1151 AAAGTGGGTT CTTGGGACC TCCGGGGTGG GGGGGTTCCA GGAGGCACGT
1201 AGGGTACCTT GCAGGGTCTT AGGAGGGAAA CCCAGGATTC CAGGAGGAT
1251 CCCAGGAACT GTGGGCACCC ATTTTCTGTG TCTCCAGCC CATTTCCTCT
1301 CCTAGTTTGT CATGGATAAT TTTTGTCTT CCCTGTGTGA TTTTGGCCAT
1351 CAAAATAAAA ATTTGAGACT CGTTAAAAA AAAAAAAAAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAAGAAA AAAAAAAAAA AAAAAA
```

BLAST Results

Entry HS806352 from database EMBL:

human STS EST192543.

Score = 1285, P = 2.5e-51, identities = 263/266

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 98 bp to 1084 bp; peptide length: 329

Category: similarity to unknown protein

Classification: no clue

1 MEVAEPSSPT EEEEEEEHS AEPRPRTSRN PEGAEDRAVG AQASVGSRSE

```

51 GEGEAASADD GSLNTSGAGP KSWQVPPAP EVQIRTPRVN CPEKVIICLD
101 LSEEMSLPKL ESFNGSKTNA LNVSQKMIEM FVRTKHKIDK SHEFALVVVN
151 DDTAWLSGLT SDPRELCSCL YDLETASCST FNLEGLFSLI QQKTELPVTE
201 NVQTIPPPYV VRTILVYSRP PCQPQFSLTE PMKKMFQCPY FFFDVVYIHN
251 GTEKEEEMS WKDMFAFMGS LDTKGTSYKY EVALAGPALE LHNCMAKLLA
301 HPLQRPCQSH ASYSLLEEED EAIEVEATV

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_78n23, frame 2

PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana, N = 1, Score = 142, P = 1.5e-07

>PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana
Length = 264

HSPs:

Score = 142 (21.3 bits), Expect = 1.5e-07, P = 1.5e-07
Identities = 56/216 (25%), Positives = 97/216 (44%)

```

Query:   93 EKVIICLDL-SEEMSLPKLESFNGSKTNALNVSQKMIEMFVRTKHKIDKSHEFALVVVND 151
          E ++IC+D+ +E M   K   NG   +   ++ I +F+ K I+ H FA   +
Sbjct:   26 EDILICIDVDAESMVEMKTTGTNGRPLIRMECVKQAILFIHNKLSINPDHRAFATLAK 85

Query:   152 DTAWLSG-LTSDPRELCSCLYDLE-TASCSTFNLEGLFSLIQQKTELPVTENVQTIPPPY 209
          AWL   TSD   + L L   S S +L LF   Q+ ++   +N
Sbjct:   86 SAAWLKKEFTSDAESAVASLRGLSGNKSSSRADLTLLFRAAAQEAQVSRQN-----R 138

Query:   210 VVRTILVYSRPPCQPQFSLTEPMKKMFQCPYFFFDVVYIHNNGTEKEEEMSWKDMF-AFM 268
          + R IL+Y R   +P   P+ +   F DV+Y+H   ++   +   +D++ + +
Sbjct:   139 IFRVILIYCRSSMRPTHEW--PLNQKL----FTLDVMYLH---DKPSPDNCPQDVYDSL 189

Query:   269 GSLD--TKGTSYKYEVALAGPALELHNCMAKLLAHPLQRPCQ 308
          +++ ++   Y +E   G A +   M+ LL HP QR Q
Sbjct:   190 DAVEHVSEYEGYIFESG-QGLARSVFKPMSMLLTHPQQRCAQ 230

```

Pedant information for DKFZphfbr2_78n23, frame 2

Report for DKFZphfbr2_78n23.2

```

[LENGTH]      329
[MW]           36560.10
[pI]           4.60
[HOMOL]        PIR:T05304 hypothetical protein F26P21.80 - Arabidopsis thaliana 7e-07
[KW]           Alpha Beta
[KW]           LOW_COMPLEXITY      9.73 %

```

```

SEQ  MEVAEPSSPTEEEEEEHSAEPRPRTRSNPEGAEDRAVGAAQASVGSRSEGEGEAASADD
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccchhhhhhhhhhhccccccccccccchhhhhhhhhhhcccccccccccccc

```

```

SEQ  GSLNTSGAGPKSWQVPPAPAEVQIRTPRVNCPKVIICLDLSEEMSLPKLESFNGSKTNA
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  LNVSQKMIEMFVRTKHKIDKSHEFALVVVNDTAWLSGLTSDPRELCSCLYDLETASCST
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ehhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhcccccccc

```

```

SEQ  FNLEGLFSLIQQKTELPVTENVQTIPPPYVVRTILVYSRPPCQPQFSLTEPMKKMFQCPY
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhheee

```

```

SEQ  FFFDVVYIHNNGTEKEEEMSWKDMFAFMGSLDTKGTSYKYEVALAGPALELHNCMAKLLA
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeeeeccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhh

```

```

SEQ  HPLQRPCQSHASYSLLEEDEAIEVEATV
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

(No Prosite data available for DKFZphfbr2_78n23.2)

(No Pfam data available for DKFZphfbr2_78n23.2)

DKFZphfbr2_7a24

group: brain derived

DKFZphfbr2_7a24 encodes a novel 142 amino acid protein with similarity to the C-terminal part of transforming growth factor-beta activated kinases.

The novel protein shows only similarity to the C-terminus of such kinases; no kinase domain is present.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to C-terminus of TGF-beta-activated kinase

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1697 bp

No poly A stretch found, no polyadenylation signal found

```
1 GGGGAGAGAG GGGTTGTGAA GGGAAGCGGA AGGGAAGGGA AGGGAGGTCC
51 CGTGGGACGC TGGGGTCTGG GGTAGAGCAG GTAGCAGCGT GCTGCCCTGA
101 CAGCTGTCTC CGCTCCTCAG ATTGTCAAGT GCTGCTATGC AGCAGGTGCA
151 GCCTGGTCTC TCACTGAGTC TCTACTCCAC AAAGGCAACG ACTGGCCAAG
201 GCAGTGGCTG GCTCTGGGTT ACACAAGTGC AGACACTCAA CTAAGTGAGC
251 TGGAAGACCC AGGAGAAGGC GGAGGCTCAG GTGCCACAT GATCAGCACA
301 GCCAGGGTAC CTGCTGACAA GCCTGTACGC ATCGCCTTTA GCCTCAATGA
351 CGCCTCAGAT GATACACCCC CTGAAGACTC CATTCCTTTG GTCTTTCCAG
401 AATTAGACCA GCAGCTACAG CCCCTGCCGC CTTGTCATGA CTCCGAGGAA
451 TCCATGGAGG TGTTACAGCA GCACTGCCAA ATAGCAGAAG AATACCTTGA
501 GGTCAAAAAG GAAATCACCC TGCTTGAGCA AAGGAAGAAG GAGCTCATTG
551 CCAAGTTAGA TCAGGCAGAA GAGGAGAAGG TGGATGCTGC TGAGCTGGTT
601 CGGGAATTCG AGGCTCTGAC GGAGGAGAAT CGGACGTTGA GGTGGGCCCA
651 GTCCTCAATGT GTGGAACAAC TGGAGAAACT TCGAATACAG TATCAGAAGA
701 GGCAGGGCTC GTCCTAACTT TAAATTTTTC AGTGTGAGCA TACGAGGCTG
751 ATGACTGCCC TGTGCTGGCC AAAAGATTTT TATTTTAAAT GAATAGTGAG
801 TCAGATCTAT TGCTTCTCTG TATTACCCAC ATGACAAC TGCTATAATGA
851 GTTTACTGCT TGCCAGCTTC TAGCTTGAGA GAAGGGATAT TTTAAATGAG
901 ATCATTAAAC TGAACATATT ACTAGTATAT GTTTTGGAG ATCAGAATTC
951 TTTTCCAAAG ATATATGTTT TTTTCTTTT TAGGAAGATA TGATCATGCT
1001 GTACAACAGG GTAGAAAATG GTAAAAATAG ACTATTGACT GACCCAGCTA
1051 AGAATCGCGG GCTGAGCAGA GTTAAACCAT GGGACAAACC CATAACATGT
1101 TCACCATAGT TTCACGTATG TGTATTTTTA AATTTATGCT CTTTAATATT
1151 TCAAATATGC TCAAATTTAA ACTGTCAGAA ACTTCTCTGC ATGTATTTAT
1201 ATTTGGCAGA GTATAAACTT TTATACTCTG ATTTTATCC TTCAATGATT
1251 GATTATACTA AGAATAAATG GTCACATATC CTAAGAGCTT CTTTATGAAA
1301 TTATTAGCAG AAACCATGTT TGAACCAAAA GCACATTTGC CAATGCTAAC
1351 TGGCTGTTGT AATAATAAAC AGATAAGGCT GCATTTGCTT CATGCCATGT
1401 GACCTCAGAG TAAACATCTC TGCCCTTGCC TGTGTGTGTT CTGGGGGAGG
1451 GGGGACATGG AAAAATATTG TTTGGACATT ACTTGGGTGA GTGCCCATGA
1501 AGACATCAGT GAACTTGTA CATTGTTTT GTTTTGGATT TAAGGAGATG
1551 TTTTAGATCA GTAACAGCTA ATAGGAATAT GCGAGTAAAT TCAGAATTGA
1601 AACAAATTCT CCTTGTCTA CCTATACCA CATTCTCTCA AATTGAAGTC
1651 TTTGTTATAT GTCCATTCTT ATTCATGTAA CTTCTTTTTC ATTAAAC
```

BLAST Results

No BLAST result

Medline entries

98130593:
Role of TAK1 and TAB1 in BMP signaling in early Xenopus development.

Peptide information for frame 1

ORF from 289 bp to 714 bp; peptide length: 142
 Category: similarity to known protein

1 MISTARVPAD KPVRIAFSLN DASDDTPPED SIPLVFPELD QQLQPLPPCH
 51 DSEESMEVFR QHCQIAEEYL EVKKEITLLE QRKKELIAKL DQAEEEKVDA
 101 AELVREFEAL TEENRTLRLA QSQCVEQLEK LRIQYQKRQG SS

BLASTP hits

Entry U92030_1 from database TREMBL:
 product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1 mRNA,
 complete cds.
 Score = 343, P = 1.3e-30, identities = 69/143, positives = 104/143

Entry AB009356_1 from database TREMBL:
 product: "TGF-beta activated kinase 1a"; Homo sapiens mRNA for
 TGF-beta activated kinase 1a, complete cds.
 Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry MMPK 1 from database TREMBL:
 product: "TAK1 (TGF-beta-activated kinase)"; Mouse mRNA for TAK1
 (TGF-beta-activated kinase), complete cds.
 Score = 339, P = 2.6e-30, identities = 67/143, positives = 104/143

Entry AB009357_1 from database TREMBL:
 product: "TGF-beta activated kinase 1b"; Homo sapiens mRNA for
 TGF-beta activated kinase 1b, complete cds.
 Score = 339, P = 3.2e-30, identities = 67/143, positives = 104/143

Entry AB009358_1 from database TREMBL:
 product: "TGF-beta activated kinase 1c"; Homo sapiens mRNA for
 TGF-beta activated kinase 1c, complete cds.
 Score = 144, P = 3.8e-09, identities = 30/67, positives = 47/67

Alert BLASTP hits for DKFZphfbr2_7a24, frame 1

PIR:JC5955 transforming growth factor-beta activated kinase (EC
 -.-.-) 1a - Human, N = 1, Score = 339, P = 3e-30

>PIR:JC5955 transforming growth factor-beta activated kinase (EC -.-.-) 1a
 - Human
 Length = 579

HSPs:

Score = 339 (50.9 bits), Expect = 3.0e-30, P = 3.0e-30
 Identities = 67/143 (46%), Positives = 104/143 (72%)

Query: 1 MISTARVPADKPVRI-AFSLNDASDDTPPEDSIPLVFPELDQQLQPLPPCHDSEESMEVF 59
 MI+T+ ++KP R ++ +D++D ++SIP+ + LD QLQPL PC +S+ESM VF
 Sbjct: 437 MITTSGPTSEKPTRSHPTPDDSTDTNGSDNSIPMAYLTLDHQLQPLAPCPNSKESMAVF 496

Query: 60 RQHCQIAEEYLEVKKEITLLEQRKKELIAKLDQAEEEKVDAAELVREFEALTEENRTLRL 119
 QHC++A+EY++V+ EI LL QRK+EL+A+LDQ E+++ + + LV+E + L +EN++L
 Sbjct: 497 EQHCKMAQEYMKVQTEIALLLQRKQELVAELDQDEKQQNTSRLVQEKKLLDENKSLST 556

Query: 120 AQSQCVEQLEKLRIQYQKRQGSS 142
 QC +QLE +R Q QKRQG+S
 Sbjct: 557 YYQCKKQLEVIRSQQQKRQGTSS 579

Pedant information for DKFZphfbr2_7a24, frame 1

Report for DKFZphfbr2_7a24.1

[LENGTH] 142
 [MW] 16377.53
 [pI] 4.64
 [HOMOL] TREMBL:U92030_1 product: "TAK1"; Xenopus laevis TGF-beta-activated kinase TAK1
 mRNA, complete cds. 6e-26
 [PROSITE] CK2_PHOSPHO_SITE 3

```

[PROSITE]      PKC_PHOSPHO_SITE      2
[PROSITE]      ASN_GLYCOSYLATION     1
[PFAM]         TNFR/NGFR cysteine-rich region
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY         7.04 %
[KW]           COILED COIL           33.10 %

```

```
SEQ      MISTARVPADKPVRIAFSLNDASDDTPPEDSIPLVFPELDQQLPPLPPCHDSEESMEVFR
SEG      .....xxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccchhhhhhccccccccchhhhhh
COLS
```

[illegible]

```

SEQ      QSQCVEQLEKLRIQYQKRQGSS
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhccc
COILS    .....

```

Prosites for DKFZphfbr2_7a24.1

PS000001	114->118	ASN_GLYCOSYLATION	PDOC000001
PS000005	4->7	PKC_PHOSPHO_SITE	PDOC000005
PS000005	116->119	PKC_PHOSPHO_SITE	PDOC000005
PS000006	18->22	CK2_PHOSPHO_SITE	PDOC000006
PS000006	26->30	CK2_PHOSPHO_SITE	PDOC000006
PS000006	77->81	CK2_PHOSPHO_SITE	PDOC000006

Pfam for DKFZphfbr2_7a24.1

HMM_NAME	TNFR/NGFR cysteine-rich region	
HMM	*CpeGtYtDWNhvpqCpIcCtrCePEMGQYmvqPCTwTQNTVC*	
	C++++ + + +Q C++ E+ +++++ T + ++	
Query	49	CHDSEESMEVF-RQH--CQIAEE--YLEVKKEITLLEQRKK 84

DKFZphfbr2_7e22

group: brain derived

DKFZphfbr2_7e22.2 encodes a novel 286 amino acid protein similar to b561 cytochromes

The new protein shows strong similarity to B561 cytochromes, but contains no heme binding site. In addition, a myc-type, helix-loop-helix dimerization domain is present. This helix-loop-helix domain mediates protein dimerization and has been found in proteins such as the myc family of cellular oncogenes, proteins involved in myogenesis and vertebrate proteins that bind specific DNA sequences in various immunoglobulin chains enhancers.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes.

strong similarity to cytochrome b561

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 4254 bp

Poly A stretch at pos. 4234, polyadenylation signal at pos. 4217

```

1 GGGGACTACC CAGAGGGCTG CCGCCGCCCTC TCCAAGTTCT TGTGGCCCCC
51 GCGGTGCGGA GTATGGGGCG CTGATGGCCA TGGAGGGCTA CCGGCGCTTC
101 CTGGCGCTGC TGGGGTCGGC ACTGCTCGTC GGCTTCTGT CGGTGATCTT
151 CGCCCTCGTC TGGGTCTCTC ACTACCGAGA GGGGCTTGGC TGGGATGGGA
201 GGGCACTAGA GTTTAACTGG CACCCAGTGC TCATGGTCAC CGGCTTCGTC
251 TTCATCCAGG GCATCGCCAT CATCGTCTAC AGACTGCCGT GGACCTGGAA
301 ATGCAGCAAG CTCCTGATGA AATCCATCCA TGCAGGGTTA AATGCAGTTG
351 CTGCCATTCT TGCAATTATC TCTGTGGTGG CCGTGTTTGA GAACCACAAT
401 GTTAACAATA TAGCCAATAT GTACAGTCTG CACAGCTGGG TTGGACTGAT
451 AGCTGTCTAT TGCTATTGTG TACAGCTCTT TTCAGGTTTT TCAGTCTTTC
501 TGCTTCCATG GGCTCCGCTT TCTCTCCGAG CATTTCTCAT GCCCATACAT
551 GTTTATTCTG GAATTGTCAT CTTTGGAACA GTGATTGCAA CAGCACTTAT
601 GGGATTGACA GAGAAACTGA TTTTTCCTCT GAGAGATCCT GCATACAGTA
651 CATTCCCGCC AGAAGGTGTT TTCGTAAATA CGCTTGGCCT TCTGATCCTG
701 GTGTTCCGGG CCCTCATTTT TTGGATAGTC ACCAGACCGC AATGGAAACG
751 TCCTAAGGAG CCAAATTCTA CCATTCTTCA TCCAAATGGA GGCACGAAC
801 AGGGAGCAAG AGGTTCCATG CCAGCCTACT CTGGCAACAA CATGGACAAA
851 TCAGATTGAG AGTTAAACAA TGAAGTAGCA GCAAGGAAAA GAAACTTAGC
901 TCTGATGAGG GCTGGGCAGA GATCTACCAT GTAAAATGTT GTAGAGATAG
951 AGCCATATAA CGTCACGTTT CAAAACTAGC TCTACAGTTT TGCTTCTCCT
1001 ATTAGCCATA TGATAATTGG GCTATGTAGT ATCAATATTT ACTTTAATCA
1051 CAAAGGATGG TTTCTTGAAA TAATTTGTAT TGATTGAGGC CTATGAACCTG
1101 ACCTGAATTG GAAAGGATGT GATTAATATA AATAATAGCA GATATAAATT
1151 GTGGTTATGT TACCTTTATC TTGTTGAGGA CCACAACATT AGCACGGTGC
1201 CTTGTGCAGA ATAGATACTC AATATGTGAA TATGTGTCTA CTAGTAGTTA
1251 ATTGGATAAA CTGGCAGCAT CCCTGGCCTG TTGTCATGCA GTCATTTCCT
1301 GTTAATTCTG GGAGACAATG ATTTCAACAAC TAGAGGGAAG CAGTCCTAAA
1351 AGTTTAAAT CCGATAAGGA ATATCTGGGA CAGGGTTTAG ATCATGACTC
1401 TACACAGATA CCATGATGAG AGTATATTAA AGAAATTTAG GAAAGCACCT
1451 GGTTCCTTTT TCCCATGCC TGCCTTCTGC TCCCTCCCCA GCTGGTTTGG
1501 GCTCAAATTG TCCCTGGAGA CTAGGGTTTA TGTTAGGGTA TTGATAGATT
1551 AGAGCAGGTG GTTGAAGAGA TCTTCTCTGG TCAGACTTGG AAGAATTTCC
1601 AAAAGTGAAG TTAGCCCCAA GACTTCCCTA GGGTTGATGT ACTTTATGAT
1651 CCAGATGCTA AACTTCTTAG AATGAAAATA TGCTTCAACA CTTAAGTAGC
1701 ATACATGAGC CTACAAACCT CAGAGAGCAC TTTTCCCCAA GTTCTTGTTC
1751 TTATTTTGA AAGTACTCAC ACAGCACTTA CTATGCTCCA AACACTCCTC
1801 TAAGCACTTT ACACATATTA GCTCATTGAG TCCCCAGACA GACGGGATGA
1851 AGTAGGTATT GTTACTGTTC CCATTTTACA GGTGAGAGAT TTGAAGCCTG
1901 GGGAGGGTAG TAACTCACCC CAAGGTGACA CGGCTCATAC ATGGTGGGAC
1951 TGAGACTCAG ATGCAGGCAG TCTGGCACCT CAGTCTGGAT TCTAACCAT
2001 TCACCTAAGT ATTTTGTGCT TGTACTACTT TGACCCACCC CTGAATAAAC
2051 CTCAATTGCT GAGATGGGGT GTAGTTATTA AAGGGATGCT TTTTACCTTT
2101 TGCTGTCTGC TGTGGCAGAT TCCCCAGATA ACCAAGGAAA AGGGGCCACC
2151 CATACCTGGA AATAGGCCAT AGGGCCCTTA CTAAGCCAA CAAGCCATGG
2201 CCTACCTTGA CACTTGTGTT ATCTTAAAT TGTTGCTTGG TAACAAAAGA
2251 TTTGGACAGG CATATCTGTA GCTTTCAAGT TAATTAATTG CAATATTTT
2301 TTCTTCAGGA TTTTAGCTGC TGAACAACCT TCAGTTTGGG GCTAAAAGAG
2351 ACCTGTCTCA TGGTCTGCCC TTCCCTGGGG CAATAGCTAG GGTCTTTCCT
2401 GATTTTTATG GAATTTTAGG GGATATTTTG AGCTTTGGGT TCTCAGTAGT

```

```

2451 GAATTGAGAC TTGGAGGTGA CTTTTCATGT TTGGAGTATC ATCTCTGTCT
2501 GGGCTCTGGG CTGACAAATT AAAACCTAGA GTAGTGCTTA TGCTGAAATG
2551 ATACTTTTCA TTTTGTGGTT GATTTTTTTG CCTTCCCTTC AATTTTAAAC
2601 TGAAGCATT TAATGTGGGT AGAACTCTA CACCAATAC ACTAAACATT
2651 TTGGTGCTTA GTGGATTCTT TTTTAGGTAA CTGGTACTTA CTCCAAAGA
2701 CTGAATACAA GCCCACTCC ATCATATCCC TTAACTTCA TGA AAAACCA
2751 TTCAAGATCC CCTTGCTGCA AACTGTCTT CTCTTCTCT ACTAAATCT
2801 ATTTCCAAAA TTGGTAATAG AGCCAGAAGG ATCCCCAGTA CCCAGCCCTC
2851 TGCCTGGCAC AAAGTGGTAG CACAATTAAA TTCAGTATGG GTGGAGCATG
2901 GTACAGTCTT GGTGCCATAG AAGGAGTAGT TGCATAGTCA CACATCATT
2951 GATAAGTTGG ATGTTCATT ACATAGAGGA ACACAAAATT CCAGGTTTT
3001 TGGAGGAAGG GATTAGATAG CGACTAAGCC GCCAGAATTG AGGTGGCCAT
3051 TCCTTTTTGT ATAGGCTAAG AAACAGGTTA TCAGTGAAAA GTTAATTATG
3101 GCCTTGGCAC TAGAATAGCA CTGTGCAAA GTATTTAAGC ACCCCCCATC
3151 TCAGCCCTTT ATTTTATCTT TCATGTGGGC TAATGTGAGG ATAATCTTAC
3201 AGATATTATA GGAATTTCTT TTCTATCTTT ATGAAAAACA CGTATATAAA
3251 ATATATCTAG AAAACCTTTG TTTGAGACTC TTATTTAATG GGCTTTTGAT
3301 TCTAATGATA ATGTACCTT TATCTTTCAA AAGCTGATAT TTCCTACCTA
3351 AGCATCTCCC GAGAAAAATA TCTCATTAAA AAGCCATAA ATAATAGGGG
3401 AGAAGAAAGC CTTAGGTATC AATCCAAAA CAGTGATTGA AATTTCCCAA
3451 AATAATTATG GCTTCTGTCA TCTCCAGAGA TAATCTGGCT TGGTTTACCC
3501 CATAATCTAA TTTCAAAAA GAAAGCTTTA TTTTAACACT CATCTGAATC
3551 AACATTAAAG CCTTTTCTCT CAAAGCGTTT ATTGAGAAAC TCAATGAAT
3601 ATACTTTTGG AATTACTGTC ATCAAAAGTG TACGGCTTCC TGTGCTGCTT
3651 GTGTCAATG GAACCTGCCC TCTAAAGCAC TTTCTTTCTT TTAATTGCGT
3701 GGTTCATGT AAGCTGTGCT GTTTAGAAAC AACATCTCAG ACTTTACAAA
3751 GAAATGACAA AGAAGGCAAT TGCACTTTT AAGGGATATC GACAAGCAGT
3801 TTCTGTTTTT TAAAGGACAA AATACAGAGT GTGTGTCATT TTTAATTAGA
3851 TTCTTTCCCC TGCTGAGTTG GAAATCCAG TGCAGCACTG ATTGACCACA
3901 GTTGCCAAAT TAAAGCACA AAGACAGAAG TAAAGCTTTA TGCTAATTTT
3951 ATTTCAATAT GATAGAAAT TTATCTTGGT ATGTCTTTT TTAGATAACT
4001 CCAGCAGGAA ACTGTAACG CTATGTCTTT AGGAAAACGT AGAAGAAAGA
4051 ACATTATTAT TCTTTAATC CTACAAGGTA CTTGAAAACC TTAAGTGAAA
4101 AAGATTCTTA TCTTTTATC TTGGCGCATT TATGGAAAAA ATATTAACTG
4151 TCCTGAATAT TTTATAATT TGTAGGAAAA ATATGCATCT ATTTTCTCT
4201 GACTTCTTTT ATATAGTAAT AAAAGTTATT TTGGAAAAA AAAAAAAA
4251 AAAA

```

BLAST Results

Entry HSG20626 from database EMBL:
human STS A005227.
Score = 860, P = 3.0e-32, identities = 176/181

Medline entries

89030633:
The structure of cytochrome b561, a secretory vesicle-specific electron transport protein.

Peptide information for frame 2

ORF from 74 bp to 931 bp; peptide length: 286
Category: strong similarity to known protein
Classification: unset

```

1 MAMEGYRRFL ALLGSALLVG FLSVIFALVW VLHYREGLGW DGSALFNWH
51 PVLMTVGFEV IQGIAIIVYR LPWTWKCSKL LMKSIHAGLN AVAAILAIIS
101 VVAVFENHNH NNIANMYSLH SWVGLIAVIC YLLQLLSGFS VLLPWAPLS
151 LRAFLMPHIV YSGIVIFGTV IATALMLTE KLIFSLRDP AYSTFPPEGVF
201 VNTLGLLILV FGALIFWIVT RPQWKRPKEP NSTILHPNGG TEQGARGSMF
251 AYSGNMMDKS DSELNNEVAA RKRNLALDEA GQRSTM

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_7e22, frame 2

SWISSPROT:C561_SHEEP CYTOCHROME B561 (CYTOCHROME B-561)., N = 1, Score

342

DKFZphfbr2_7j4

group: brain derived

DKFZphfbr2_7j4 encodes a novel 233 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit

Sequenced by GBF

Locus: unknown

Insert length: 1050 bp

Poly A stretch at pos. 1027, polyadenylation signal at pos. 1007

```

1  GGGGACACAA AGGGGTGGTC ACCCTGCCCT CACCTTGACC TGTAAGTTGC
51 CTAGGACAGT GGCCTGGTCC CAGGGGCTGT TGTGGGGAGT TGAAGAACAC
101 CCTGGCCTCC TCCATCATGT CGGCCAAGAG GGCAGAATTG AAGAAAACAC
151 ATCTGTGCAA GAACTACAAG GCAGTTTGCC TGGAAATTGA GCCAGAGCCG
201 ACCAAAACAT TTGATTACAA AGCAGTTAAA CAAGAAGGGC GGTTTACCAA
251 AGCAGGAGTG ACACAGGACC TAAAGAATGA ACTCAGGGAA GTGAGAGAAG
301 AGCTCAAGGA GAAAATGGAG GAGATAAAAC AGATAAAGGA TCTAATGGAC
351 AAGGATTTTG ATAAACTTCA CGAATTTGTG GAAATTATGA AGGAAATGCA
401 GAAAGATATG GATGAGAAGA TGGACATTTT AATAAATACA CAGAAGAACT
451 ATAAGCTTCC CTTAGAAGA GCACCAAAGG AGCAGCAGGA ACTCAGGCTG
501 ATGGGAAAGA CTCACAGAGA ACCACAGCTC AGGCCCAAGA AAATGGATGG
551 AGCCAGTGGG GTCAATGGAG CACCCTGTGC TCTTCACAAG AAGACGATGG
601 CACCACAAAA AACAAAACAG GGCTCACTGG ATCCCCTTCA TCACTGTGGG
651 ACCTGCTGCG AGAAATGTTT GTTGTGTGCT CTAAAGAACA ACTACAATCG
701 GGGGAACATT CCTTCAGAGG CCTCAGGCCT TTACAAAGGT GGAGAGGAGC
751 CAGTGACCAC CCAACCTTCT GTGGGCCACG CTGTGCCTGC CCCAAGTCC
801 CAGACTGAGG GAAGGTGAAG CTTAACTGCC AGCTTGAAT GAGAGTAAAG
851 AAGATACAGA GCAACAGTG TTTCAGAAAC TGTCTGCCCC TGGGTGTGAT
901 TCTTTGGCTT CAATTTGAAG GAGGAGGAAT GATGGGATTT CATATTTTAT
951 TTCACACCAG TTCCTCCTTG TTTCACTCTT TTGCTAAGCT GGCTGCTTCT
1001 ACCATCTAAT AAATAATTGG CCAAGTTAAA AAAAAAAAAA AAAAAAAAAA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 117 bp to 815 bp; peptide length: 233
 Category: putative protein

```

1  MSAKRAELKK THLCKNYKAV CLELKPEPTK TFDYKAVKQE GRFTKAGVTQ
51  DLKNELEVR EELKEKMEEI KQIKDLMDKD FDKLHEFVEI MKEMQKDMDE
101 KMDILINTQK NYKLPLRRAP KEQELRLMG KTHREPQLRP KMDGASGVN
151 GAPCALHKKK MAPQKTKQGS LDPLHHCCTC CEKCLLCALK NNYNRGNIPS
201 EASGLYKGGE EPVTTQPSVG HAVPAPKSQT EGR

```

BLASTP hits

Entry JC2223 from database PIR:
 major surface glycoprotein 3 - *Pneumocystis carinii* (fragment)
 Score = 109, P = 3.5e-04, identities = 41/136, positives = 67/136

Alert BLASTP hits for DKFZphfbr2_7j4, frame 3

TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence., N = 1, Score = 109, P = 0.00024

>TREMBLNEW:PCP115C_1 product: "P115C"; Pneumocystis carinii mRNA for P115C, partial sequence.
Length = 196

HSPs:

Score = 109 (16.4 bits), Expect = 2.4e-04, P = 2.4e-04
Identities = 41/134 (30%), Positives = 67/134 (50%)

```
Query:   14 CKN-YKAVCLELKPPTKTFDYKAVKQEGRTKA-GVTQDLKNELREVREELKEKMEEIK 71
          CK  K  C  ELK      + K VK+  TK  G  ++LK++++  E  KE++E  K
Sbjct:   22 CKTELKKYCEELKEADGLKVNDK-VKEICDDTKRDGKCKELKDKVKKELETFKEELE--K 78

Query:   72 QIKDLMDKDFDKLHEFVEIMKEMQKMDDEKMDILINTQKNYKPLRRAPKEQQLRLMGK 131
          +KD+ D++ +K  E  +++E  D D K + +  + YKL +R  E  LR +GK
Sbjct:   79 ALKDIDENCEKYEEKCILLEETNHD-DVKKNCKVLREGCYKLRKRVA-EDLLLRALGK 136

Query:   132 THREPQLRPKKMDGAS 147
          +  +  K  D  S
Sbjct:   137 DVKNGECEKMKDVCS 152
```

Pedant information for DKFZphfbr2_7j4, frame 3

Report for DKFZphfbr2_7j4.3

```
[LENGTH]      233
[MW]           26533.95
[pI]           9.18
[PROSITE]      MYRISTYL           3
[PROSITE]      CK2_PHOSPHO_SITE   3
[PROSITE]      PKC_PHOSPHO_SITE   3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY     14.59 %
[KW]           COILED_COIL       13.73 %
```

```
SEQ  MSAKRAELKKTHLCKNYKAVCLELKPPTKTFDYKAVKQEGRTKAGVTQDLKNELREVR
SEG  .....XXXXXXXXX
PRD  ccchhhhhhhhhccchhhhhhhccccccccccccccccccccccccchhhhhhhhh
COILS .....CCCCCCCCCCCC
```

```
SEQ  EELKEKMEEIKQIKDLMDKDFDKLHEFVEIMKEMQKMDDEKMDILINTQKNYKPLRRAP
SEG  .....XXXXXXXXX
PRD  hhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccc
COILS CCCCCCCCCCCCCCCCCCCCCC.....
```

```
SEQ  KEQQLRLMGKTHREPQLRPKKMDGASGVNGAPCALHKKTMAPQKTKQGSLDPLHHCCTC
SEG  .....
PRD  hhhhhhhhhccccccccccccccccccccccccchhhhhcccccccccccccccccc
COILS .....cccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ  CEKCLLCALKNNYNRGNIPSEASGLYKGGEPTVTQPSVGHAVPAPKSQTTEGR
SEG  .....
PRD  chhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....cccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphfbr2_7j4.3

PS00005	2->5	PKC_PHOSPHO_SITE	PDOC00005
PS00005	108->111	PKC_PHOSPHO_SITE	PDOC00005
PS00005	132->135	PKC_PHOSPHO_SITE	PDOC00005
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	179->183	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00008	151->157	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_7j4.3)

DKFZphfbr2_82c20

group: transmembrane protein

DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with very weak similarity to C. elegans cosmid D1007.

The novel protein contains 7 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans D1007.5 ;
membrane regions: 7
Summary DKFZphfbr2_82c20 encodes a novel 492 amino acid protein with
similarity to a hypothetical C.elegans protein.

similarity to C.elegans D1007.5

complete cDNA (Bp 1-100 GC ritch), complete cds,
potential start at Bp 128 matches Kozak consensus PyNNatgG,
EST hits, localisation? primer B of STS doesn't match perfect!
TRANSMEMBRANE 7

Sequenced by DKFZ

Locus: /map="109.9 cR from top of Chr1 linkage group"???

Insert length: 1804 bp
Poly A stretch at pos. 1794, no polyadenylation signal found

```
1 CGGCGGGAGC GCGCGGCTGA TACCCGGGAC TGGGCTGCGG CGGTTAGTCC
51 TCTCCCGGCC GCCGTGCGCT CCGACATATT GCTCGCAGGA GCTGCGGCGG
101 CGAAGCGGAG AGCACCGGGG GGAGGAGATG GGAGGACGAA GAGGTCCCAA
151 CAGGACATCT TACTGTCGAA ATCCGCTCTG TGAGCCGGGA TCCTCGGGGG
201 GCTCTAGTGG AAGCCACACT TCCAGTGCAT CGGTGACCAG TGTTCTGTTCC
251 CGCACCAGGA GCAGTTCTGG AACAGGCCTC TCCAGCCCTC CTCTGGCCAC
301 CCAAACTGTT GTGCCCTCTAC AGCACTGCAA GATCCCCGAG CTGCCAGTCC
351 AGGCCAGCAT TCTGTTTGAG TTGCAGCTCT TCTTCTGCCA GCTCATAGCA
401 CTCTTCGTCC ACTACATCAA CATCTACAAG ACAGTGTGGT GGTATCCACC
451 TTCCACCCA CCTCCACACA CCTCCCTGAA CTTCCATCTG ATCGACTTCA
501 ACTTGCTGAT GGTGACCACC ATCGTTCTGG GCCGCGGCTT CATTGGGTCC
551 ATCGTGAAGG AGGCCTCTCA GAGGGGGAAG GTCTCCCTCT TTCGCTCCAT
601 CCGTGTGTTT CCACTCTGCT TCACCGTTCT CACGGCAACA GGCTGGAGTC
651 TGTGCCGATC CCTCATCCAC CTCTTCAGGA CTTACTCTCT CCTGAACCTC
701 CTGTTCTCTT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCAGCTGAA
751 TTGCGACCTC CGCAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC
801 CCCGGGAGGC GGTGAGTGGC CTGGCAAAGA GCCGGGACTA CCTCCTGACA
851 CTGCGGGAGA CGTGAAGACA GCACACAAGA CAGCTGTATG GCCCGGACGC
901 CATGCCCCACC CATGCCTGCT GCCTGTCAAC CAGCCTCATC CGCAGTGAGG
951 TGGAGTTCTT CAAGATGGAC TTCAACTGGC GCATGAAGGA AGTGCTCGTC
1001 AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCCTG TCTGGTTCGT
1051 GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGAA CTCTTCCTGC
1101 TGGTGTCATC CAGCACCTCC GTGATCCTCA TGCAGCACCT GCTGCCTGCC
1151 AGCTACTGTG ACCTGCTGCA CAAGGCGGCC GCCCATCTGG GCTGTTGGCA
1201 GAAGGTGGAC CCAGCGCTGT GCTCCAACGT GCTGCAGCAC CCGTGGACTG
1251 AAGAATGCAT GTGGCCGCGC GCGCTGCTGG TGAAGCACAG CAAGAACGTC
1301 TACAAAGCCG TAGGCCACTA CAACGTGGCT ATCCCTCTG ACCTCTCCCA
1351 CTTCGGCTTC CATTCTTTT TCAGCAAACC TCTGCGGATC CTCAACATCC
1401 TCCTGCTGCT GGAGGGCGCT GTCATTGTCT ATCAGCTGTA CTCCCTAATG
1451 TCCTCTGAAA AGTGGCACCA GACCATCTCG CTGGCCCTCA TCCTCTTCTCAG
1501 CAACTACTAT GCCTTCTTCA AGCTGCTCCG GGACCGCTTG GTATTGGGCA
1551 AGGCCTACTC ATACTCTGCT AGCCCCCAGA GAGACCTGGA CCACCGTTTC
1601 TCCTGAGCCC TGGGGTCACC TCAGGGACAG CGTCCAGGCT TCAGCCAAGG
1651 GCTCCTTGCC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA
1701 GACAAAAAAA TCCACCAGAG CTTTGTATTT TTGTTACGTA CTGTTTCTTT
1751 GATAATTGAT GTGATAAGGA AAAAAGTCCT ATTTTATATC TCCCAAAAAA
1801 AAAA
```

BLAST Results

Entry HS285343 from database EMBL:
human STS WI-17488.

Score = 1225, P = 1.3e-50, identities = 263/281

Medline entries

No Medline entry

Peptide information for frame 2

```
1  MGGRRGPNRT  SYCRNP LCEP  GSSGGSSGSH  TSSASVTSVR  SRTRSSSGTG
51 LSSPPLATQT  VVPLQHKIP  ELPVQASILF  ELQLFFCQLI  ALFVHYINIY
101 KTVWVYPPSH  PPSHTSLNFH  LIDFNLLMVT  TIVLGRRFIG  SIVKEASQRG
151 KVSILFRSILL  FLTRFTVLTA  TGWSLCRSLI  HLFRTYSFLN  LLFLCYPFGM
201 YIPFLQLNCD  LRKTSLFNFM  ASMGPRAVS  GLAKSRDYLL  TLRETWKQHT
251 RQLYGPDAMP  THACCLSPSL  IRSEVEFLKM  DFNWRMKEVL  VSSMLSAYYV
301 AFVPVWFVKN  THYYDKRWSC  ELFLVLSIST  SVILMQHLLP  ASYCDLLHKA
351 AAHLGCWQKV  DPALCSNVLQ  HPWTECMWP  QGVLVKHSKN  VYKAVGHYNV
401 AIPSDVSHFR  FHFFFSKPLR  ILNILLLEG  AVIVYQLYSL  MSSEKWHQTI
451 SLALILFSNY  YAFFKLLRDR  LVLGKAYSYS  ASPQRDLDR  FS
```

ORF from 128 bp to 1603 bp; peptide length: 492
Category: similarity to unknown protein
Prosites motifs: LEUCINE ZIPPER (210-232)
LEUCINE_ZIPPER (210-232)

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfbr2_82c20, frame 2

TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid
D1007., N = 2, Score = 247, P = 4.6e-29

>TREMBL:CEAF3151_8 gene: "D1007.5"; Caenorhabditis elegans cosmid D1007.
Length = 512

HSPs:

Score = 247 (37.1 bits), Expect = 4.6e-29, Sum P(2) = 4.6e-29
Identities = 58/204 (28%), Positives = 102/204 (50%)

```
Query: 291 VSSMLSAYYVAFVPVWFVKNTHYYDKRWSCEFLVLSISTSVILMQHLLPASVCDLLHKA 350
      +S ML +V F + ++ W C+L ++V ++ + + +L P +Y DLLH+A
Sbjct: 299 LSIMLPCIFVPFKTSQGIPQKILINEVWECQLAIVVGLTAFSLYVAYLSPLNYDLLHRA 358

Query: 351 AAHLGCWQKVD-PAL---CSNVLQHPWTEECMWPGQVLVKHSKN-VYKAVGHYNV---- 400
      A HLG W +++ P + + PW+E C++ G V+ Y+A ++
Sbjct: 359 AIHLGSWHQIEGPRIGHTGSMSSAPTPWSEFCLYNDGETVQMPDGRGYRAKSSNSIRTVA 418

Query: 401 AIPSDVSHFRFHFFFSKPLRILNILLLEGAVIVYQLYSLMSSEKWHQTIISLALILFSNY 460
      A P H F KP ++NI+ E +I Q + L+ + W ++ L++F+NY
Sbjct: 419 AHPSSRHNTFFKVLKPNLNIMCSFEFLIFIQFWMVLVLTNDWQHIVTFVLLMFANY 478

Query: 461 YAFFKLLRDRVLVKGAYSYSASPQRDL 487
      F KL +D+++L + Y S Q DL
Sbjct: 479 LLFAKLKDKIILSRIYEPS---QEDL 502
```

Score = 178 (26.7 bits), Expect = 4.3e-21, Sum P(2) = 4.3e-21
Identities = 50/179 (27%), Positives = 90/179 (50%)

```
Query: 262 HACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYVAFVPVWFV--KNTHYYDKR-- 317
      H C SP+ IR E++ L D R+K+ + + + +A+ +P EV K + ++
Sbjct: 262 HMCSDSPAQIREIEIQVLIDDLVLRVKKSIAGVSTAFSLSIMLPCIFVPFKTSQGIPQKIL 321

Query: 318 ----WSCFLVLSISTSVILMQHLLPASVCDLLHKAHAAHLGCWQKVD-PAL---CSNV 368
      W C+L ++V ++ + + +L P +Y DLLH+AA HLG W +++ P + +
Sbjct: 322 INEWECQLAIVVGLTAFSLYVAYLSPLNYDLLHRAAIHLGSWHQIEGPRIGHTGSMSS 381

Query: 369 LQHPWTEECMWPGQVLVKHSKN-VYKAVGHYNV-AIPSDVSHFRFHFFFSKPLRILNILL 426
      PW+E C++ G V+ Y+A ++ + + R + FF K LR N L+
Sbjct: 382 APTWSEFCLYNDGETVQMPDGRGYRAKSSNSIRTVAHPSSRHNTFF-KVLRKPNLI 440
```

```

SEQ      MGGRRGPNRTSYCRNP LCEPGSSGGSGSHTSSASVTSVRSTRSSSGTGLSSPPLATQT
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD      cccccccccccccccccccccccccccccccccccceecccccccccccccccccccc
MEM      .....

SEQ      VVPLOHQCKIPELPVQASILFELQLFFCQLIALFVHYINIKTVWVYPPSHPSHTSLNFH
SEG      .....
PRD      eeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhheeeccccccccceeeee
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMM

SEQ      LIDFNLLMVTITVLGRRFISIVKEASRGKVS LFRSILLFLTRFTVL TATGWSLCRSLI
SEG      .....
PRD      eehhhhhhhhhhhhhheeehhhhhhhhc cchhhhhhhhhhhhhhhhhhhhhc cchhhhhhh
MEM      MMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      HLFRTYSFLNLLFLCYPFMGYIPFLQLNC DL RKTSLFNHMASMGPREAVSGLAKSRDYLL
SEG      .....
PRD      hhhhhhhhhheeeeeccccceeeeccccchhhhhhhhhc cchhhhhhhhhhhhhhhhh
MEM      .....

SEQ      TLRETWKQHTRQLYGP DAMPTHACCLSPSLIRSEVEFLKMDFNWRMKEVLVSSMLSAYYV
SEG      .....
PRD      hhhhhhhhhhhhc cccccccccccccccchhhhhhhhhhhhhhhhhhhhhc chhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMM

SEQ      AFVPVWFVKNTHYDKRSCELFLLVSISTSVILMQHLLPAS YCDLLHKAA AHLGCWKQV
SEG      .....
PRD      heeeeeeeccccchhhhhhhhhc cchhhhhhhhhhhc cchhhhhhhhhhhhhhhhhcc
MEM      MMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      DPALCSNVLQHPWTEECMW PQGV LVKHSKNVYKAVGHYNVAIPSDVS HFRHF FFSKPLR
SEG      .....xx
PRD      cccccccccccccceec cccceeeec cccceec cccccc cccccceeeeeeccch
MEM      .....MMMMMMMMMM

SEQ      ILNILLLEGA VIVYQLYSLMSSEKWHQT ISALILFSNYA FF KLLRDRLVL GKAYSYS
SEG      xxxxxxxx
PRD      hhhhhhhhhheeeehhhhhc cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc ccc
MEM      MMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMM

```



```

SEQ   ASPQRDLDRHFS
SEG   .....
PRD   ccchhhhhcccc
MEM   .....

```

Prosites for DKFZphfbr2_82c20.2

PS00001	8->12	ASN_GLYCOSYLATION	PDOC00001
PS00002	47->51	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	212->216	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	316->320	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	443->446	PKC_PHOSPHO_SITE	PDOC00005
PS00006	241->245	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	28->34	MYRISTYL	PDOC00008
PS00008	48->54	MYRISTYL	PDOC00008
PS00008	231->237	MYRISTYL	PDOC00008
PS00009	2->6	AMIDATION	PDOC00009
PS00009	134->138	AMIDATION	PDOC00009
PS00029	168->190	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphfbr2_82c20.2)

DKFZphfbr2_82e17

group: transmembrane protein

DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with very weak similarity to C. elegans cosmid R01B10.

The novel protein contains 6 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

similarity to C.elegans "R01B10.5" ;
membrane regions: 6
Summary DKFZphfbr2_82e17 encodes a novel 311 amino acid protein with
similarity to a hypothetical C.elegans protein.

similarity to C.elegans "R01B10.5"

complete cDNA, EST HS763158 extends the sequence, complete cds, EST
hits
six potential transmembrane domains

Sequenced by DKFZ

Locus: /map="779_C_?; 818_A_1; 877_C_1; 734_C_12; 760_E_11; 171.7 cR from top of Chr14 linkage
group"

Insert length: 1618 bp

Poly A stretch at pos. 1608, polyadenylation signal at pos. 1588

```

1 CTGATCTAGT GCTTCTCGAA AAAAACCTTC AGGCGGCCCA TGGCTGTCGA
51 TATTCAACCA GCATGCCTTG GACTTTATTG TGGGAAGACC CTATTATTTA
101 AAAATGGCTC AACTGAAATA TATGGAGAAT GTGGGGTATG CCCAAGAGGA
151 CAGAGAACGA ATGCACAGAA ATATTGTCTG CCTTGCACAG AATCTCCTGA
201 ACTTTATGAT TGGCTCTATC TTGGATTATG GGCAATGCTT CCTCTGGTTT
251 TACATTGGTT CTTCATTGAA TGGTACTCGG GGAAAAAGAG TTCCAGCGCA
301 CTTTTCACAC ACATCACTGC ATTATTGAA TGCAGCATGG CAGCTATTAT
351 CACCTTACTT GTGAGTGATC CAGTTGGTGT TCTTTATATT CGTTCATGTC
401 GAGTATTGAT GCTTCTGAC TGGTACACGA TGCTTTACAA CCCAAGTCCA
451 GATTACGTTA CCACAGTACA CTGTACTCAT GAAGCCGCTT ACCCACTATA
501 TACCATTGTA TTTATCTATT ACGCATTCTG CTTGGTATTA ATGATGCTGC
551 TCCGACCTCT TCTGGTGAAG AAGATTGCAT GTGGGTAGG GAAATCTGAT
601 CGATTATAAA GTATTATGTC TGCACTTTAC TTCTTCCCAA TTTTAACCGT
651 GCTTCAGGCA GTTGGTGGAG GCCTTTTATA TTACGCCTTC CCATACATTA
701 TATTAGTGTT ATCTTTGGTT ACTCTGGCTG TGTACATGTC TGCTTCTGAA
751 ATAGAGAACT GCTATGATCT TCTGGTCAGA AAGAAAAAGC TTATTGTTCT
801 CTTCAGCCAC TGGTACTTTC ATGCCTATGG AATAATCTCC ATTTCCAGAG
851 TGGATAAACT TGAGCAAGAT TTGCCCTTTT TGGCTTTGGT ACCTACACCA
901 GCCCTTTTTT ACTTGTTCAC TGCAAAATTT ACCGAACCTT CAAGGATACT
951 CTCAGAAAGG GCCAATGGAC ACTGAGTGTA GACATGTGAA ATGCCAAAAA
1001 CCTGAGAAGT GCTCCTAATA AAAAAGTAAA TCAATCTTAA CAGTGTATGA
1051 GAACATTCTT ATCATATATG GGAACAAGAT TGTCAGTATA TCTTAATGTT
1101 TGGGTTTGTC TTTGTTTGTG TTATGGTTAG ACTTACAGAC TTGGAAATG
1151 CAAAACCTCG TAATACTCTG TTACACAGGG TAATATTATC TGCTACACTG
1201 GAAGGCCGCT AGGAAGCCCT TGCTTCTCTC AACAGTTTCA CTGTTCTTTA
1251 GGGCAAAATC ATGTTTCTGT GTACCTAGCA ATGTGTTCCC ATTTTATTAA
1301 GAAAAGCTTT AACACGTGTA ATCTGCAGTC CTTAACAGTG GCGTAATTGT
1351 ACGTACCTGT TGTGTTTCAG TTTGTTTTTC ACCTATAATG AATTGTAAAA
1401 ACAAAACATAC TTGTGGGGTC TGATAGCAAA CATAGAAATG ATGTATATTG
1451 TTTTGTGTTA TCTATTTATT TTCAATCAATA CAGTATTTTG ATGTATTGCA
1501 AAAATAGATA ATAATTTATA TAACAGGTTT TCTGTTTATA GATTGGTTCA
1551 AGATTGTGTT GGATTATTGT TCCTGTAAAG AAAACAATAA TAAAAAGCTT
1601 ACCTACATAA AAAAAAAA

```

BLAST Results

Entry HS981146 from database EMBL:
human STS WI-6253.
Length = 208
Minus Strand HSPs:
Score = 1040 (156.0 bits), Expect = 1.9e-40, P = 1.9e-40

Identities = 208/208 (100%), Positives = 208/208 (100%), Strand = Minus
/ Plus

Entry HSG20716 from database EMBL:

human STS A006D06.

Length = 195

Minus Strand HSPs:

Score = 975 (146.3 bits), Expect = 1.8e-37, P = 1.8e-37

Identities = 195/195 (100%), Positives = 195/195 (100%), Strand = Minus
/ Plus

Medline entries

No Medline entry

Peptide information for frame 1

```
1 MAVDIQPACL GLYCGKTLF KNGSTEIYGE CGVCPRGQRT NAQKYCQPCT
51 ESPELYDWLY LGFMAMLPV LHWWFFIEWYS GKKSSSALFQ HITALFECM
101 AAIITLLVSD PVGVLYIRSC RVLMLSDWYT MLYNPSPDYV TTVHCTHEAV
151 YPLYTIVFIY YAFCLVLMML LRPLLVKKIA CGLGKSDRFK SIYAALYFFP
201 ILTVLQAVGG GLLYYAFPIY ILVLSLVTLA VYMSASEIEN CYDLLVRKKR
251 LILVLFHWLL HAYGIISISR VDKLEQDLPL LALVPTPALF YLFTAKFTEP
301 SRILSEGANG H
```

ORF from 40 bp to 972 bp; peptide length: 311
Category: similarity to unknown protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82e17, frame 1

TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid
R01B10., N = 1, Score = 399, P = 1.4e-36

>TREMBL:AF068718_5 gene: "R01B10.5"; Caenorhabditis elegans cosmid R01B10.
Length = 670

HSPs:

Score = 399 (59.9 bits), Expect = 1.4e-36, P = 1.4e-36

Identities = 95/280 (33%), Positives = 152/280 (54%)

```
Query: 2 AVDIQPACLGLYCGKTLFKN-----GSTEIYGECGVCPRGQRTNAQKYCQPC 49
      A IQP+CLG +CG+T+L N GST + CG C G R NA C+ C
Sbjct: 292 ASTIQPSCLG-FCGRTVLVGNYSIEDVEATTTAAGSTSL-SRCGPCSFGYRNNAMSICESC 349

Query: 50 TESPELYDWLYLGFMAMLPVLHWWFFIEWYSGKKSSSALFQ---HITALFECSMIAITL 106
      + YDW+YL F+A+LPL+LH FI + K + ++ ++ + E +A +I +
Sbjct: 350 DTPLQPYDWMYLLFIALPLLLHMQFIR-IARKYCRTRYEVSEYLCVILENVIACVIAV 408

Query: 107 LVSDPVGVLVIRSCRVLMLSDWYTMLYNPSPDYVTVHCTHEAVYPLYTIVFIYYAFCLV 166
      L+ P ++ C + +WY YNP Y T+ CT+E V+PLY+I FI++ +
Sbjct: 409 LIYPPRFTFFLNGCSKTDIKWYPACYNPRIGYTKTMRCTYEVVFPYISITFIHHLILIG 468

Query: 167 LMLLLRPLLVKKIACGLGKSDRFKSIYAALYFFPILTVLQAVGGGLLYYAFPIIILVLSL 226
      +++LR L + L K+ K YAA+ PIL V+ AV G+++Y FPYI+L+ SL
Sbjct: 469 SILVLRSTLYCVL---LYKTYNGKPFYAAIVSVPI LAVIHAVLSGVVFYTFPYILLIGSL 525

Query: 227 VTLAVYMSASEIENCYDLLVR---KKRLIVLFSHWLLHAYGIISI 268
      + +++ ++VR LI L L+ ++G+I+I
Sbjct: 526 WAMCFHLALEGKRPLKEMIVRIATSPTHLIFLSITMLMLSFGVIAI 571
```

Pedant information for DKFZphfbr2_82e17, frame 1

Report for DKFZphfbr2_82e17.1

Prosites for DKFZphfbr2_82e17.1

(No Pfam data available for DKFZphfbr2_82e17.1)

DKFZphfbr2_82e4

group: signal transduction

DKFZphfbr2_82e4 encodes a novel 473 amino acid protein with strong similarity to the calmodulin-binding proteins.

The novel protein is similar to human and rat Ca²⁺/calmodulin-dependent protein kinase (EC 2.7.1.123), rat calmodulin-binding protein, calmodulin binding protein kinase of Fugu rufes and Rattus norvegicus calcium/calmodulin-dependent protein kinase I. Calmodulin is the archetype of the family of calcium-modulated proteins of which nearly 20 members have been found. Calmodulin is involved in regulation of growth and cell cycle as well as in signal transduction and the synthesis and release of neurotransmitters. The novel protein seems to be involved in calmodulin-mediated pathways in human neuronal cells.

The new protein can find clinical application in modulating/blocking calmodulin-mediated pathways in human neuronal cells.

strong similarity to calmodulin-binding proteins

complete cDNA, complete cds, EST hits
splice variant in comparison to rat I56542
ESTs HS2254543/HS1141907 define splice variant
see also DKFZphfbr2_82g20 unspliced form

Sequenced by DKFZ

Locus: /map="200.5 cR from top of Chr3 linkage group"

Insert length: 2923 bp

Poly A stretch at pos. 2913, polyadenylation signal at pos. 2890

```
1 ATGCTGGAGG TTCGCTAGCC GAAGCGGCTG CATCTGGCGC CGCGTCTGCC
51 CCGCGTGCCT GGAGCGGATT CTGCCCGCCG TCCCGGAGC CCTCGGCGCC
101 CCCTGAGCC CGCGATCACT TCCTCCCTGT GACCAACCGG CGCTGCAGGT
151 TAGAGCCTGG CAATGCCGTT TGGGTGTGTG ACTCTGGGTG ACAAGAAGAA
201 CTATAACCAAG CCATCGGAGG TGAATGACAG ATATGATTTC GGACAGGTCA
251 TCAAGACTGA GGAGTTTGTG GAAATCTTCC GGGCCAAGGA CAAGACGACA
301 GGCAAGCTGC ACACCTGCAA GAAGTTCCAG AAGCGGGACG GCCGCAAGGT
351 GCGGAAAGCT GCCAAGAAGC AGATAGGCAT CCTCAAGATG GTGAAGCATC
401 CCAACATCCT ACAGCTGGTG GATGTGTTTG TGACCCGCAA GGAGTACTTT
451 ATCTTCTCTGG AGCTGGCCAC GGGGAGGGAG GTGTTTGACT GGATCCTGGA
501 CCAGGGCTAC TACTCGGAGC GAGACACAAG CAACGTGGTA CGGCAAGTCC
551 TGGAGGCCCT GGCCATTATG CACTCACTCA AGATCGTGCA CAGGAATCTC
601 AAGCTGGAGA ACCTGGTTTA CTACAACCGG CTGAAGAAGT CGAAGATTGT
651 CATCAGTGAC TTCCATCTGG CTAAGCTAGA AAATGGCCTC ATCAAGGAGC
701 CCTGTGGGAC CCCCAGATAT CTGGGCAACC CACCTTTCTA TGAGGAGGTG
751 GAAGAAGATG ATTATGAGAA CCATGATAAG AATCTCTTCC GCAAGATCCT
801 GGCTGCTGAC TATGAGTTTG ACTCTCCATA TTGGGATGAT ATTTGCGAGG
851 CAGCCAAAGA CCTGGTCACA AGGCTGATGG AGGTGGAGCA AGACCAGCGG
901 ATCACTGCAG AAGAGGCCAT CTCCCATGAG TGGATTCTCT GCAATGCTGC
951 TTCTGATAAG AACATCAAGG ATGGTGTCTG TGCCCAAGAT GAAAGAAGAT
1001 TTGCCAGGGC CAAGTGAAGG AAGGCTGTCC GAGTGACCAC CCTCATGAAA
1051 CCGCTCCGGG CACCAGAGCA GTCCAGCAGC GCTGCAGCCC AGTCGGCCTC
1101 AGCCACAGAC ACTGCCACCC CCGGGGCTGC AGGTGGGGCC ACAGCTGCAG
1151 CTGCGAGTGG AGCTACCTCA GCCCCTGAGG GTGATGCTGC TCGTGCTGCA
1201 AAGAGTGATA ATGTGGCCCC CGCAGACCGT AGTGCCACCC CAGCCACAGA
1251 TGGAAAGTGC ACCCCAGCCA CTGATGGCAG TGTACCCCA GCCACCGATG
1301 GAAGCATCAC TCCAGCCACT GATGGGAGTG TCACCCAGC CACTGACAGG
1351 AGCGCTACTC CAGCCACTGA TGGGAGAGCC ACACCAGCCA CAGAAGAGAG
1401 CACTGTGCCC ACCACCCAAA GCAGTGCCAT GCTGGCCACC AAGGCAGCTG
1451 CCACCCCTGA GCCGGCTATG GCCAGCCGG ACAGCACAGC CCCAGAGGGC
1501 GCCACAGGCC AGGCTCCACC CTCTAGTAAA GGGGAAGAGG CTGCTGGTTA
1551 TGCCCAAGGAG TCTCAAAGGG AGGAGGCCAG CTGAGTAGGC AGCCTGGTGA
1601 GGGGGGGCAG GGGATGGGCA GGAGGGTGGG AGAGTGGATG AGGGGCTTCT
1651 CACTGTACAT AGAGTCACTG GCATGATGCC CTCGCTCCCC CATGCCCCCA
1701 CATCCCAAGT GGGCATAACT AGGGGTACAG GGAGAGCAGT CTCGTCTCCT
1751 GTGTGTATGT GTGTGAGTGG TGGGAGGGCC AGTGGCAGGG CCGGCCCCAG
1801 CCCCTGCATG GATTCTTTGT GGCTTTTCTG TCTTTTGCTA GCTTCACCAT
1851 TTTCTGTTCG TTGTGGGATG CTGCTCTAGG GATACTCAGG GGGCTCCTGC
1901 TCTCCTTCCC CTTCCTTCTT TGCTCACCA TTCCCTTAGG CAGGCCCTGC
1951 AGGTCCACCA CTCTCCAGG CCTAAACTT GGGCGGCCTT GCCCTGAGAG
2001 CTGGTCCCTC AGCGAGGCC TGTCAGCGGT CTTAGGCTCC TGCACATGAA
2051 GGTGTGTGCC TGTGGTGTGT GGGCTGCTCT AGGAGCAGAT ACAGGCTGGT
2101 ATAGAGGATG CAGAAAGGTA GGGCAGTATG TTTAAGTCCA GACTTGGCAC
2151 ATGGCTAGGG ATACTGCTCA CTAGCTGTGG AGTCTCTCAG GAGTGGAGAG
2201 AATGAGTAGG AGGGCAGAA CTCCATTTT TGTCCTTCCT AAGACCTAGT
```

```

2251 TATTTGTGTT ATTTCCCTGCC TTTCCGAGTC CTGCAGTGGG CTGCCCTGTA
2301 CCCTGAACCT CATGAGCCTC TAAGGGAAAG GAGGAACAAT TAGGACGTGG
2351 CAATGAGACC TGGCAGGGCA GAGTACAAGC CCAGCACCCA GTGTCCCAGC
2401 CTTACTGGGT CCTTACCCTG GGCCAAACAG GGAGGGCTGA TACCTCCTTG
2451 CTCTTCCTAG ATGCCCACCT CCTACAATCT CAGCCACAA GTCTCTCCA
2501 CCCTAGGGGG CTTGCTGCAT GGCAATAACT CATAATCTGA TTTGGAGGTT
2551 TGCCCTTTAC AGGGGCAGAT TTTCTGCTCA GTTCAACAAT GAAATGAAGA
2601 GGAACCTCCCT CTTTCTACAG CTCACTTCTA TCAGAGGCC AGGTGCCTCA
2651 GAGCCACATT GAGTTGCTTT TTCTGGGATG AGGAAGTAGG GTTAACTCC
2701 CCAGTTTCCT GAGGGAGGCT CCTGACAGGT GCCCTTTGTC AGACCCTACC
2751 ACAGCCTGGA TAGGCAGCCA CATTGGTCCT CGCCCTTGCT CGGCACTCCG
2801 TGGTGGTCCT GCCCTTCTCC CTGCATGCCT GTGGGTCTGC TCTGGTGTGT
2851 GAAGGTCGGT GGGTTAACTG TGTGCCTACT GAACCTGGCA AATAAACATC
2901 ACCCTGCAAA GCCAAAAAA AAA

```

BLAST Results

```

-----
Entry HS452352 from database EMBL:
human STS WI-15318.
Length = 350
Minus Strand HSPs:
Score = 1547 (232.1 bits), Expect = 5.2e-63, P = 5.2e-63
Identities = 331/348 (95%), Positives = 331/348 (95%), Strand = Minus /
P1

```

Medline entries

```

-----
94110847:
J Neurosci 1994 Jan;14(1):1-13
IG5: a calmodulin-binding, vesicle-associated, protein
kinase-like protein enriched in forebrain neurites.
Godbout M, Erlander MG, Hasel KW, Danielson PE, Wong KK, Battenberg EL,
Foye PE,
Bloom FE, Sutcliffe JG

```

Peptide information for frame 1

```

-----
1 MPFGCVTLGD KKNYNQPSEV TDRYDLGQVI KTEEFCEIFR AKDKTTGKLN
51 TCKKFKQRDG RKVRKAAKNE IGILKMKVHP NILQLVDVFN TRKEYFIFLE
101 LATGREVFDFW ILDQGYYSER DTSNVVRQVL EAVAYLHSLK IVHRNLKLEN
151 LVYYNRLKNS KIVISDFHLA KLENGLIKEP CGTPEYLGNP PFYEEVEEDD
201 YENHDKNLFR KILAGDYEDF SPYWDDISQA AKDLVTRLME VEQDQRITAE
251 EATSHIEWISG NAASDKNIKD GVCAQIEKNF ARAKWKKAVER VTTLMKRLRA
301 PEQSSTAAQA SASATDTATP GAAGGATAAA ASGATSAPEG DAARAAKSDN
351 VAPADRSATP ATDGSATPAT DGSVTPATDG SITPATDGSV TPATDRSATP
401 ATDGRATPAT EESTVPTQS SAMLATKAAA TPEPAMAQPD STAPEGATGQ
451 APPSSKGEEA AGYAQESQRE EAS

```

ORF from 163 bp to 1581 bp; peptide length: 473
Category: strong similarity to known protein

BLASTP hits

```

Entry S50193 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - rat
Length = 374
Score = 371 (130.6 bits), Expect = 2.2e-66, Sum P(2) = 2.2e-66
Identities = 74/176 (42%), Positives = 115/176 (65%)

```

```

Entry S57347 from database PIR:
Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.123) I - human
Length = 370
Score = 369 (129.9 bits), Expect = 4.6e-66, Sum P(2) = 4.6e-66
Identities = 74/176 (42%), Positives = 114/176 (64%)

```

Alert BLASTP hits for DKFZphfbr2_82e4, frame 1

PIR:I56542 calmodulin-binding protein - rat, N = 2, Score = 1246, P = 4e-228

TREMBLNEW:FRU010348_3 product: "calmodulin binding protein kinase";
Fugu rubripes UBE1-like gene, PRGFR2 gene and gene encoding calmodulin
binding protein kinase, clone 168J21, N = 2, Score = 846, P = 2.6e-139

TREMBL:RNPRKI_1 product: "protein kinase I"; Rattus norvegicus
calcium/calmodulin-dependent protein kinase I mRNA, complete cds., N =
2, Score = 364, P = 5.1e-63

>PIR:I56542 calmodulin-binding protein - rat
Length = 504

HSPs:

Score = 1246 (186.9 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 255/289 (88%), Positives = 259/289 (89%)

Query: 188 GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSFYWDDISQAAKDLVTRLMEVEQDQRI 247
GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSFYWDDISQAAKDLVTRLMEVEQDQRI
Sbjct: 216 GNPPFYEEVEEDDYENHDKNLFKILAGDYEFDSFYWDDISQAAKDLVTRLMEVEQDQRI 275

Query: 248 TAEAAISHEWISGNAASDKNIKDGVCAGIEKNFARAKWKKAVRVTTLMKRLRAPEQSSTA 307
TAEAAISHEWISGNAASDKNIKDGVCAGIEKNFARAKWKKAVRVTTLMKRLRAPEQS TA
Sbjct: 276 TAEAAISHEWISGNAASDKNIKDGVCAGIEKNFARAKWKKAVRVTTLMKRLRAPEQSGTA 335

Query: 308 AAQSASATDTATPGAAGGATAAAASGATSAPE-----GDAARAASDNVAPADRSAT 359
A +D ATPGAAGGA AAAA GA A GDA AAKSD++A ADRSAT
Sbjct: 336 AT-----SDAATPGAAGGAVAAAAGGAAPASGASATVGTGGDAGCAAKSDDMASADRSAT 390

Query: 360 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQ 419
PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVP Q
Sbjct: 391 PATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPAAQ 450

Query: 420 SSAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS 473
SSA A KAAATPEPA+AQPDSTA EGATGQAPPSSKGEEA G AQESQR E S
Sbjct: 451 SSAAPAAKAAATPEPAVAQPDSTALEGATGQAPPSSKGEEATGCAQESQRVETS 504

Score = 978 (146.7 bits), Expect = 4.0e-228, Sum P(2) = 4.0e-228
Identities = 186/187 (99%), Positives = 187/187 (100%)

Query: 1 MPFGCVTLGDKKNYNQPSSEVTDRLYDLGQVITKEEFCEIFRAKDKTTGKLHTCKKFQKRDG 60
MPFGCVTLGDKKNYNQPSSEVTDRLYDLGQV+KTEEFCEIFRAKDKTTGKLHTCKKFQKRDG
Sbjct: 1 MPFGCVTLGDKKNYNQPSSEVTDRLYDLGQVVKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG 60

Query: 61 RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVDFWILDQGYYSER 120
RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVDFWILDQGYYSER
Sbjct: 61 RKVRKAANEIGILKMVKHPNQLQVDFVTRKEYFIFLELATGREVDFWILDQGYYSER 120

Query: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180
DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
Sbjct: 121 DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP 180

Query: 181 CGTPEYL 187
CGTPEYL
Sbjct: 181 CGTPEYL 187

Pedant information for DKFZphfbr2_82e4, frame 1

Report for DKFZphfbr2_82e4.1

[LENGTH] 473
[MW] 51208.89
[pI] 5.30
[HOMOL] PIR:I56542 calmodulin-binding protein - rat 0.0
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-30
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 2e-26
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YDL101c] 8e-26
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 5e-24
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 7e-23
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
7e-23
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 1e-21
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 1e-21

[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w] 3e-19
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-16
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 3e-16
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 1e-15
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 3e-14
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YCR073c] 6e-11
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNR031c] 8e-11
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YJL095w] 2e-09
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 1e-08
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c] 7e-08
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YFL033c] 1e-07
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 5e-07
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 8e-07
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c] 5e-06
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-05
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w] 1e-05
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c] 8e-05
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR523c] 2e-04
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-04
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins
[BLOCKS] BL00939F
[SCOP] digol_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus] 3e-62
[SCOP] dlwfc_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens)] 5e-59
[SCOP] dikoa_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis] 1e-75
[SCOP] dikoba_ 5.1.1.1.6 Twitchin, kinase domain [california sea har] 1e-72
[SCOP] dlphk_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 4e-65
[SCOP] dlirk_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens)] 2e-56
[SCOP] dlapme_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu)] 4e-71
[SCOP] dlfgka_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Homo)] 1e-50
[SCOP] dlydre_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo)] 3e-70
[SCOP] dlfrm_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Homo)] 5e-49
[SCOP] dlcdkb_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su)] 2e-72
[SCOP] d2hcka_3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [human (Huma)] 5e-46
[SCOP] dlcsn_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe] 9e-42
[SCOP] dljsua_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens)] 1e-56
[SCOP] dlckia_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus)] 9e-52
[EC] 2.7.1.38 Phosphorylase kinase 3e-29
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 8e-66
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 2e-17
[EC] 2.7.1.117 Myosin-light-chain kinase 2e-38
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 2e-17
[EC] 2.7.1.37 Protein kinase 6e-28
[PIRKW] phosphotransferase 8e-66
[PIRKW] nucleus 2e-24
[PIRKW] transferase 8e-30
[PIRKW] calcium 2e-27
[PIRKW] duplication 4e-19
[PIRKW] tandem repeat 2e-31
[PIRKW] phorbol ester binding 1e-16
[PIRKW] zinc 1e-16
[PIRKW] cell cycle control 2e-20
[PIRKW] serine/threonine-specific protein kinase 8e-66
[PIRKW] phospholipid binding 1e-16
[PIRKW] autophosphorylation 8e-66
[PIRKW] brain 1e-14
[PIRKW] heterotetramer 2e-16
[PIRKW] polymer 3e-29
[PIRKW] mitosis 2e-20
[PIRKW] magnesium 7e-22
[PIRKW] ATP 8e-66
[PIRKW] alternative initiators 1e-29

[PIRKW] phosphoprotein 8e-66
 [PIRKW] apoptosis 2e-31
 [PIRKW] glycoprotein 4e-19
 [PIRKW] skeletal muscle 3e-28
 [PIRKW] protein kinase 2e-28
 [PIRKW] testis 3e-28
 [PIRKW] signal transduction 1e-21
 [PIRKW] cAMP binding 1e-16
 [PIRKW] purine nucleotide binding 5e-25
 [PIRKW] structural protein 4e-19
 [PIRKW] calcium binding 3e-45
 [PIRKW] alternative splicing 3e-45
 [PIRKW] P-loop 5e-25
 [PIRKW] lipoprotein 2e-16
 [PIRKW] cardiac muscle 4e-19
 [PIRKW] muscle 3e-28
 [PIRKW] myristylation 2e-16
 [PIRKW] EF hand 5e-29
 [PIRKW] cell division 2e-38
 [PIRKW] calmodulin binding 8e-66
 [PIRKW] smooth muscle 7e-31
 [SUPFAM] fibronectin type III repeat homology 7e-31
 [SUPFAM] immunoglobulin homology 7e-31
 [SUPFAM] ribosomal protein S6 kinase II 3e-26
 [SUPFAM] calcium-dependent protein kinase 5e-29
 [SUPFAM] AMP-activated protein kinase 7e-22
 [SUPFAM] protein kinase akt 1e-14
 [SUPFAM] protein kinase SPK1 3e-20
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-36
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-45
 [SUPFAM] calmodulin repeat homology 5e-29
 [SUPFAM] protein kinase DUN1 2e-24
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-14
 [SUPFAM] death-associated protein kinase 2e-31
 [SUPFAM] myosin-light-chain kinase, nonmuscle 1e-29
 [SUPFAM] pleckstrin repeat homology 1e-14
 [SUPFAM] ankyrin repeat homology 2e-31
 [SUPFAM] protein kinase homology 8e-66
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-36
 [SUPFAM] twitchin 1e-18
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-16
 [SUPFAM] titin 4e-19
 [SUPFAM] protein kinase cdrl 2e-20
 [SUPFAM] kinase-related transforming protein 2e-38
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 8e-66
 [SUPFAM] kinase interaction domain homology 2e-24
 [SUPFAM] protein kinase C mu 1e-16
 [PROSITE] AMIDATION 1
 [PROSITE] MYRISTYL 3
 [PROSITE] CK2_PHOSPHO_SITE 10
 [PROSITE] TYR_PHOSPHO_SITE 2
 [PROSITE] PKC_PHOSPHO_SITE 11
 [PFAM] Eukaryotic protein kinase domain
 [KW] All_Alpha
 [KW] 3D
 [KW] LOW_COMPLEXITY 7.40 %

SEQ MPFGCVTLGDKKNYNQPSVETDRYDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDG
 SEG
 1a06-CEETTTGGGCEEEEEECBCGGGGGEEEEETTTTCEEEEEEEEC---

 SEQ RKVRKAAKEIGILKMVKHPNQLQVDVFVTRKEYFIFLELATGREVFDWILDQGYYSER
 SEG
 1a06- -----HHHHHHHHHCCTTTBCCEEEEETTEEEEECCCEHHHHHHHTTTTBHH

 SEQ DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAKLENGLIKEP
 SEG
 1a06- HHHHHHHHHHHHHHHHHHCCTTTTTTTEEECCCTTTTCEEECCCTTTTCHHHHHHCCC

 SEQ CGTPEYLGPNPPFYEEVEEDDYENHDKNLFRKILAGDYEFDSPYWDDISQAADLVTRLME
 SEG
 1a06- HHHHHHHHCCTTTTTT-----THHHHHHHHHCCCCCTTTTTTCHHHHHHHHHHCT

 SEQ VEQDQRITAEAAISHWISGNAASDKNIKDGVCQAIKFNFAKAKKKAVRVTTLMKRLRA
 SEG
 1a06- TTGGGCCCHHHHHHTTTTTTCCCCCBHHHHHHHHHHHCCTTTTTTBTHHHHHHHHC..

 SEQ PEQSSTAAQASATDTATPGAAGGATAAASGATSAPEGDAARAASDNVAPADRSATP
 SEG
 1a06-

```

SEQ      ATDGSATPATDGSVTPATDGSITPATDGSVTPATDRSATPATDGRATPATEESTVPTTQS
SEG      .....
1a06-    .....

SEQ      SAMLATKAAATPEPAMAQPDSTAPEGATGQAPPSSKGEEAAGYAQESQREEAS
SEG      .....
1a06-    .....

```

Prosites for DKFZphfbr2_82e4.1

```

PS00005      21->24      PKC_PHOSPHO_SITE      PDOC00005
PS00005      46->49      PKC_PHOSPHO_SITE      PDOC00005
PS00005      51->54      PKC_PHOSPHO_SITE      PDOC00005
PS00005      91->94      PKC_PHOSPHO_SITE      PDOC00005
PS00005      103->106     PKC_PHOSPHO_SITE      PDOC00005
PS00005      118->121     PKC_PHOSPHO_SITE      PDOC00005
PS00005      138->141     PKC_PHOSPHO_SITE      PDOC00005
PS00005      264->267     PKC_PHOSPHO_SITE      PDOC00005
PS00005      394->397     PKC_PHOSPHO_SITE      PDOC00005
PS00005      454->457     PKC_PHOSPHO_SITE      PDOC00005
PS00005      467->470     PKC_PHOSPHO_SITE      PDOC00005
PS00006      7->11       CK2_PHOSPHO_SITE      PDOC00006
PS00006      91->95       CK2_PHOSPHO_SITE      PDOC00006
PS00006      103->107     CK2_PHOSPHO_SITE      PDOC00006
PS00006      118->122     CK2_PHOSPHO_SITE      PDOC00006
PS00006      248->252     CK2_PHOSPHO_SITE      PDOC00006
PS00006      313->317     CK2_PHOSPHO_SITE      PDOC00006
PS00006      336->340     CK2_PHOSPHO_SITE      PDOC00006
PS00006      442->446     CK2_PHOSPHO_SITE      PDOC00006
PS00006      455->459     CK2_PHOSPHO_SITE      PDOC00006
PS00006      467->471     CK2_PHOSPHO_SITE      PDOC00006
PS00007      456->464     TYR_PHOSPHO_SITE      PDOC00007
PS00007      127->136     TYR_PHOSPHO_SITE      PDOC00007
PS00008      260->266     MYRISTYL              PDOC00008
PS00008      321->327     MYRISTYL              PDOC00008
PS00008      324->330     MYRISTYL              PDOC00008
PS00009      59->63      AMIDATION             PDOC00009

```

Pfam for DKFZphfbr2_82e4.1

```

HMM_NAME      Eukaryotic protein kinase domain

HMM            *YeigRiIGeGsFGtVYkCiWr.TGeIvAIKIIkkrms.....FlREIq
               Y +G++I   F ++++++ TG++   K++ KR+   + +EI
Query          24  YDLGQVIKTEEFCEIFRAKDKTTGKLHTCKKFQKRDGRKVRKAANEIG      72

HMM            IMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEwe
               I+++++HPNI+++ D+F  + +++ + +E++ G + FD+I ++G++SE++
Query          73  ILKMVKHPNILLVDVFEV-TRKEYFIFLELATGREVFDWILDQGYYSERD    121

HMM            IrfIMYQILrGMeYLHSMgIIHRDLKPENILIDeN...gqIKicDFGLAR
               ++++Q+L++++YLHS +I+HR LK EN+ + ++   I I+DF LA+
Query          122 TSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYYNRLKNSKIVISDFHLAK    171

HMM            qMnnYerMttfCGTPWY*
               + N  ++ +  CGTP+Y
Query          172 LEN--GLIKEPCGTPEY      186

HMM            *GepPFYd.....dnMemImrIiqrfrprfWpnCSeElyDFMr
               G PPFY+   + +++I+++++F +P+W+ +S  ++D+++
Query          188 GNPPFYEEVEEDDYENHOKNLFKILAGDYEFDSPYWDNISQAADLVT      236

HMM            wCWnyDPekRPTFrQILnHPWF*
               +++++  ++R+T+++++ H W+
Query          237 RLMEVEQDQRITAEAEISHEWI      258

```

DKFZphfbr2_82g14

group: transmembrane protein

DKFZphfbr2_82g14 encodes a novel 208 amino acid proline-rich protein without similarity to known proteins.

The protein contains one transmembrane domain.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of brain-specific genes and as a new marker for neuronal cells.

unknown prolin rich protein
membrane regions: 1

Summary DKFZphfbr2_82g14 encodes a novel 208 amino acid protein.

unknown prolin rich protein

complete cDNA, complete cds, EST hits
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="26.2 cR from top of Chr16 linkage group"

Insert length: 2059 bp

Poly A stretch at pos. 2049, polyadenylation signal at pos. 2024

```
1 AGAAGTGCGA CTGCCAGCTG CCGAGGCGTT CGGTCCTGCT GTTGGGGCCG
51 CTGCCCCAGG GCTGCGGGGA CGCTCCCGGA GCCCTGCCTG TCCCTGTGCC
101 ATCCAGGCCA GCAGCTGAAG GAGCCTCACC TGCCCTCCCTT CTCTGAGTAG
151 CACGGATTGT AGGAGAAGCA GCGAAGATGT CCAGCGAGCC TCCCCCTCCT
201 TATCCTGGGG GCCCCACAGC CCCACTTCTG GAAGAGAAAA GTGGAGCCCC
251 GCCCACCCCA GGCCGTTCTT CCCCAGCTGT GATGCAGCCC CCTCCAGGCA
301 TGCCACTGCC CCCTGCGGAC ATTGGCCCCC CACCCATATGA GCCGCCGGGT
351 CACCCAATGC CCCAGCCTGG CTTCATCCCA CCACACATGA GTGCAGATGG
401 CACCTACATG CCTCCGGGTT TCTACCCCTCC TCCAGGCCCC CACCCACCCA
451 TGGGCTACTA CCCCCAGGG CCCTACACGC CAGGGCCCTA CCCTGGCCCT
501 GGGGGCCACA CAGCCACAGT CCTGGTCCCT TCAGGAGCTG CCACCACGGT
551 GACAGTGTCT CAGGGAGAGA TCTTTGAGGG AGCGCCTGTG CAGACGGTGT
601 GTCCCACTG CCAGCAGGCC ATCGCCACCA AGATCTCCTA CGAGATTGGC
651 TTGATGAATT TCGTGCTGGG TTTCTTCTGT TGCTTCATGG GATGTGATCT
701 GGGTGCTGCT CTGATCCCTT GCCTCATCAA TGAATTCAAG GATGTGACGC
751 ACACATGCCC CAGCTGCAAA GCCTACATCT ACACGTACAA GCGCCTGTGC
801 TAACGGAGCT GGGACTCGGG ACTCCCCCGC CTGTCACTCT GGCCCCCTGT
851 GCTTTGCTCC CTGCGCTCAG TGGTCACTTT CCCGCTCCCA CTTGGGGCTG
901 GGAGCCGTGC CACCATCCCC TAGAAGTCCT GTCCCTCTTA CCCTGCCCTA
951 CCTGAGCCGC TGACTCTTCT GGCAAAAATT CTGTTGGGAT TTAAGGCCAA
1001 GGGTCAGTGG GTGGCAGGGG GCTGGCAATG AGCTTGTGTG TTGTTGGTCT
1051 GCTTGGTGTG TGTGATCGGG AAGATAAGCT GGGAGGGGTC TCCTGCTGGG
1101 GTCCGTATGC CTCTGTTTCC AAACAAGGTA CAGGTTCACT CCAGACTCTT
1151 TCCCCCTGGG ACCAACAGCA GCCAGAGCAG TTAGCCAGTT AGTCCCAGG
1201 CCTGTGGCCA CAGGCGTTTC TGACCTGCTG GGCCGAGAAT GGGTAAGTTG
1251 TCTGGAGTCA GGTGGGCCCA CGTAGGACAG GGTACAAAAG CCTGGGTTTG
1301 TTTCTGGGTA CTTTGCGCCT CTGGGGTGCT AGAGGTGGGG CATGGTGGCT
1351 GGAAGTAAAA CTGCCAACTC TGGCCCTCAG AACTCTCAGG TATAGAAGCC
1401 CAGGATGTCT AATACCCCTG CCCAGTGCCC GAGAGCTGCC TGGTGTACAG
1451 TAGAGAGGAC ACTGTACCTG GGTGAATGAT CAGACCCTGG TAGCTAAGAA
1501 GGAACCTTGT CCTTTGAGTC AGTGTGCAGA CCCCCTTTCA GGCCATGCCT
1551 CTGTGAACCC TGTATTGCTG GGGCCGGAAG GAGCCCTGA GCCTAGCCCC
1601 TTCCCGTCTG CCCTGTGTCC TCACTGCGTG TGGGTATGAC CTCTGCCTGG
1651 TGGCTGGTGT ATCCCAACTG GGCAAGAGAT GGCAGAGGGT CCCCCTGTGT
1701 GGTGCGCTTG GATGTGCAGA GCCTTCTCCA TGGATTTTCT TCCCTGTAAG
1751 TGCCGGGCCC CCCACCCAG CTGACAGGCT GTTGCTGTGC CTGCTCACAC
1801 CTGCTCCTGC AGGCACACTG GGCTAGGGAC GAGGAAGGAG CAGCCACAAG
1851 TGGTAGAACT GCCTTGGTGG ACACCAAGCT CGCCCTGTCT TTATTTCCTG
1901 AATGTTTGT GAACTTGCTC ACCTGGACCA CTGTATCCTG CCACTGTCTG
1951 TCCTGGTCTC GCACTGCCAC TGCATGGCCT CCTGTCACTG TGAATCGTGG
2001 CCCAGTCTCA GTTTGTAGTT TCTCATTAAA TTGGCCCTTT CACTCCCCCA
2051 AAAAAAAA
```

BLAST Results

Entry HS727347 from database EMBL:
human STS WI-16589.
Length = 275
Plus Strand HSPs:
Score = 1365 (204.8 bits), Expect = 3.0e-55, P = 3.0e-55
Identities = 275/276 (99%), Positives = 275/276 (99%), Strand = Plus /
p1

Medline entries

No Medline entry

Peptide information for frame 3

1 MSSEPPPPYP GGPTAPLLEE KSGAPPTPGR SSPAVMQPP GMPLPPADIG
51 PPPYEPGPH MFQPGFIPPH MSADGYTMP GFYPPPGPH PMGYPPPGYP
101 TPGYPGPGG HTATVLVPSG AATTVTVLQG EIFEGAPVQT VCPHCQQAIA
151 TKISYEIGLM NEVLGFFCCF MGC DLGCLLI PCLINDFKDV THTPCSKAY
201 IYTYKRLC

ORF from 177 bp to 800 bp; peptide length: 208
Category: similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82g14, frame 3

PIR:S57447 HPBRII-7 protein - human, N = 1, Score = 206, P = 8.4e-16

PIR:A47655 spliceosome-associated protein SAP 62 - human, N = 1, Score = 198. P = 4.3e-15

>PIR:S57447 HPBRII-7 protein - human
Length = 551

HSPs :

Score = 206 (30.9 bits), Expect = .8.4e-16, P = 8.4e-16
Identities = 57/115 (49%), Positives = 62/115 (53%)

Query: 5 PPPPYGGPTAPLLEEKSGAPPTGRSSPAVMQPPGMPLPPADIGPP-----PYEP--- 56
 PPPP+P G T P G P PG P PPPG LPP GPP P P
 Sbjct: 226 PPPFPAGQTPP--RPPLGPPGPPGPPG---PPPGQVLPPPLAGPPNRGRPPPPVLF 279

Query: 57 PGHPMPQP--GFIPPHMSADGTYPMPGEFYPPPGHPHPM-GYYPP-GPYTPGPYPGPGGH 111
PG P QP G +PP G P PG+ PPPGP PP G PP GP+ P P PGP G
Sbjct: 280 PGQPFQGPLPLGPP-----GPPPPVPGYGPMPGPPPPPGGPPPPPGFPFPPRP-PGLPLG 333

```
Query: 112 TATVLVP 118
      T+ P
Sbjct: 334 PLTLAPP 340
```

Score = 177 (26.6 bits), Expect = 1.1e-12, P = 1.1e-12
Identities = 55/120 (45%), Positives = 61/120 (50%)

Query: 5 P P P P P G G P T A P - - L L E E K S G A P P T P G - R S S P A V M - - - Q P - - - P P G M P L P P A D I G P P P Y E 55
P P P P G P P + L P P G R P V + Q P P P P L P P G P P P
Sbjct: 244 P G P P G P P G P P P P G O V L P P P L A G P P N R G D R P P P P V L F P G Q P F G Q P P L G P L P P - - - G P P P - P 299

Query: 56 PPGHMPQPFGFIPPHMSADGTYMPPGPFYPP--PGP-HPPMGYYPPGPYTPGPYPG---PG 109
PG+ P PG PP G PPG +PP PGP PP+ PP P+ PGP PG P
Sbjct: 300 VPGYG-PPPGPPPPQO---GPPPPPGPFPPRPPGGLGPELTLAPP-PHLGPPPGAPPPA 354

```
Query:   110 GHTATVLVP 118
          H      P
Sbjct:   355 PHVNPAFFP 363
```

Score = 168 (25.2 bits), Expect = 1.1e-11, P = 1.1e-11
Identities = 47/118 (39%), Positives = 51/118 (43%)

Query: 5 PPPPYPG-GPTAPLLEEKSGAPPTPGRSSPAVMQP--PPGMPLPPADI-GPPPYEPPGHP 60

Sbjct: 296 PPPPGG GP + G PP PG P P PP PP + GPPP PP P
 Query: 61 MPQPGFIPPHMSADGTYMPPGFYPPPGPHPPMGYYPPGPTYTPGYPGPGGHTATVLPVPSG 120
 P F PP ++ MP P P P G PP PY G Y PG T P
 Sbjct: 356 HVNPAFFPPPTNSG---MPTSDSRGPPPTDPYGR-PP-PYDRGDYGPFGREMDTARTPLS 410
 Query: 121 AA 122
 A
 Sbjct: 411 EA 412

Score = 156 (23.4 bits), Expect = 2.1e-10, P = 2.1e-10
 Identities = 44/103 (42%), Positives = 50/103 (48%)

Query: 6 PPPYGGGPTAPLLEEKSGAPPT-PGRSSPAVMQPPPGMPLPPADIGPPPYEPPGHMPQP 64
 P PGG P G PP P +P +PP G P PP GPPP PG +P P
 Sbjct: 208 PGAVPGGDRFPGPAGPGGPPPPFFAGQTPP--RPPLGPPGPPGPPGPP--PGQVLPPP 262

Query: 65 GFIPPHMSADGTYMPPGFYPPPGPHPPMGYYPPGPTYTP---GPYPGP 108
 PP+ D PP +P P PP+G PPGP P GP PGP
 Sbjct: 263 LAGPPNRG-DRP-PPVLFPGQFPGQPLGLPLPGPPPPVPGYGPVPG 309

Score = 121 (18.2 bits), Expect = 5.2e-05, P = 5.2e-05
 Identities = 40/90 (44%), Positives = 45/90 (50%)

Query: 23 GAPPTPGRSSPAVMQPP-PGMPLPPAD-IGPP-PYEPPGHMPMPQG-FIPPHMSADGTYM 78
 G PG + P PP P PP +GPP P PPG P P PG +PP ++
 Sbjct: 213 GDRFPFGPAGPGGPPPPFFAGQTPPRPPLGPPGPPGPPG-P-PPPGQVLPPPLAG----- 265

Query: 79 PP--GFYPPPG---PHPPMGYYPPGPTYTPGYPG-PG 109
 PP G PPP P P G P GP PGP P PG
 Sbjct: 266 PPNRGDRPPPPVLFPGQFPGQPLGLPLPGPPPPVPG 302

Pedant information for DKFZphfbr2_82g14, frame 3

Report for DKFZphfbr2_82g14.3

[LENGTH] 208
 [MW] 21862.47
 [pI] 5.55
 [PROSITE] MYRISTYL .3
 [PROSITE] PKC_PHOSPHO_SITE 2
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 39.90 %

SEQ MSSEPPPPYGGGPTAPLLEEKSGAPPTPGRSSPAVMQPPPGMPLPPADIGPPPYEPPGHP
 SEGXX
 PRD cccccccccccccchhhhhcc
 MEM

SEQ MPQPGFIPPHMSADGTYMPPGFYPPPGPHPPMGYYPPGPTYTPGYPGPGGHTATVLPVPSG
 SEGXX
 PRD ccc
 MEM

SEQ AATTVTVLQGEIFEGAPVQTVCPHCQQAIAIKISYEIGLMNFVLGFFCCFMGCDLGCCLI
 SEG
 PRD cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhcecccccccccccccccccc
 MEMMMMMMMMMMMMMMM

SEQ PCLINDFKDVTHTCPSCKAYIYTYKRLC
 SEG
 PRD eeeeecc
 MEM MMMM.....

Prosite for DKFZphfbr2_82g14.3

PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	172->178	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_82g14.3)

DKFZphfbr2_82i17

group: signal transduction

DKFZphfbr2_82i17 encodes a novel 334 amino acid protein with similarity to the plasma membrane substrate for the cAMP-dependent protein kinase.

The novel protein is a transmembrane protein with strong similarity to the phospholemman protein, a membrane substrate for the cAMP-dependent protein kinase. It seems to serve as a chloride channel or as a chloride-channel regulator.

The new protein can find application in modulating/blocking cAMP-dependent protein kinase-dependent pathways.

similarity to plasma membrane substrate for cAMP-dependent protein kinase

complete cDNA, complete cds, EST hits
potential start at Bp 31 matches Kozak consensus PyNnatgG
might be a SODIUM/POTASSIUM-TRANSPORTING ATPASE
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="11: 920_E_12; 786_(A,H)_11; (797,802)_(E,H)_7"

Insert length: 1647 bp

Poly A stretch at pos. 1637, polyadenylation signal at pos. 1615

```
1 AGTCTCGGAG GGGACCGGCT GTGCAGACGC CATGGAGTTG GTGCTGGTCT
51 TCCTCTGCAG CCTGTGGGCC CCCATGGTCC TGGCCAGTGC AGCTGAAAAG
101 GAGAAGGAAA TGGACCCCTT TCATTATGAT TACCAGACCC TGAGGATTGG
151 GGGACTGGTG TTCGTGTGGG TTCTCTTCTC GGTGGGATC CTCCTTATCC
201 TAAGTCGCAG GTGCAAGTGC AGTTTCAATC AGAAGCCCCG GGGCCAGGGA
251 GATGAGGAAG CCCAGGTGGA GAACCTCATC ACCGCCAATG CAACAGAGCC
301 CCAGAAAGCA GAGAACTGAA GTGCAGCCAT CAGGTGGAAG CCTCTGGAAC
351 CTGAGGCGGC TGCTTGAACC TTTGGATGCA AATGTGATG CTTAAGAAAA
401 CCGGCCACTT CAGCAACAGC CCTTCCCCA GGAGAAGCCA AGAACTTGTG
451 TGTCCCCAC CCTATCCCTT CTAACACCAT TCCTCCACCT GATGATGCAA
501 CTAACACTTG CCTCCCGCT GCAGCCTGTG GTCCTGCCA CCTCCCGTGA
551 TGTGTGTGTG TGTGTGTGTG TGTGTGACTG TGTGTGTTG CTAAGTGTG
601 TCCTTGTGGC TACTTGTGTT TGGATGGTAT TGTGTTGTT AGTGAAGTGT
651 GGAATCGCTT TCCAGGCGAG GGGCTGAGCC ACACGGCCAT CTGCTCCTCC
701 CTGCCCCCGT GGCCCTCCAT CACCTTCTGC TCCTAGGAGG CTGCTTGTG
751 CCGGAGACCA GCCCCCTCCC CTGATTTAGG GATGCGTAGG GTAAGAGCAC
801 GGGCAGTGGT CTTCACTCGT CTTGGGACCT GGAAGGTTT GCAGCACTTT
851 GTCATCATTC TTCATGGACT CCTTCACTC CTTTAACAAA AACCTTGCTT
901 CCTTATCCCA CTGATCCCA GTCTGAAGGT CTCTTAGCAA CTGGAGATAC
951 AAAGCAAGGA GCTGGTGAGC CCAGCGTTGA CGTCAGGAGC GCTATGCCCT
1001 TCCGTGGTTA ATTTCTTCCC AGGGGCTTCC ACAGAGAGTC CCCATCTGCC
1051 CCGCCCTTTC ACAGAGCGCC CGGGGATTCC AGGCCAGGG CTTCTACTCT
1101 GCGCCCTGGG AATGTGTCCC CTGCATATCT TCTCAGCAAT AACTCCATGG
1151 GCTCTGGGAC CCTACCCCTT CCAACCTTCC CTGCTTCTGA GACTTCAATC
1201 TACAGCCAG CTCATCCAGA TGCAGACTAC AGTCCCTGCA ATTGGGTCTC
1251 TGGCAGGCAA TAGTTGAAGG ACTTCTGTGT CCGTTGGGGC CAGCACACCG
1301 GGATGGATGG AGGGAGAGCA GAGGCCTTGT CTTCTCTGCC TACGTCCCTT
1351 TAGATGGGCA GCAGAGGCAA CTCCCGCATC CTTTGTCTGT CCTGTGAGTG
1401 GTCAGAGCGG TGAGCGAGGT GGGTTGGAGA CTCAGCAGGC TCCGTGCAGC
1451 CCTTGGGAAC AGTGAGAGGT TGAAGGTCAT AACGAGAGTG GGAATCAAC
1501 CCAGATCCCG CCCCTCCTGT CCTCTGTGTT CCGCGGAAA CCAACCAAC
1551 CGTGCGCTGT GACCCATTGC TGTCTCTGT ATCGTGACCT ATCCTCAACA
1601 ACAACAGAAA AAAGGAATAA AATATCCTTT GTTCTCTAAA AAAAAAA
```

BLAST Results

Entry HS31455 from database EMBL:
human STS WI-2739.
Length = 103
Minus Strand HSPs:
Score = 487 (73.1 bits), Expect = 4.4e-14, P = 4.4e-14
Identities = 101/104 (97%), Positives = 101/104 (97%), Strand = Minus /
Plus
frame shift in primer binding site

Medline entries

91250422:
Purification and complete sequence determination of the major plasma
membrane substrate
for CAMP-dependent protein kinase and protein kinase C in myocardium.

95091702:
Protein kinase C and cyclic AMP-dependent protein kinase phosphorylate
phospholemman,
an insulin and adrenaline-regulated membrane phosphoprotein, at
specific sites in the
carboxy terminal domain.

95138184:
Mat-8, a novel phospholemman-like protein expressed in human breast
tumors, induces a
chloride conductance in Xenopus oocytes.

Peptide information for frame 2

1 MELVLVFLCS LLAPMVLASA AEKEKEMDPF HYDYQTLRIG GLVFAVVLFS
51 VGILLILSRR CKCSFNQKPR APGDEEAQVE NLITANATEP QKAEN

ORF from 32 bp to 316 bp; peptide length: 95
Category: strong similarity to known protein

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i17, frame 2

SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 196, P =
1.2e-15

TREMBL:AF091390.1 product: "phospholemman precursor"; Mus musculus
phospholemman precursor, gene, complete cds., N = 1, Score = 187, P =
1.1e-14

PIR:A40533 CAMP-dependent protein kinase major membrane substrate
precursor - dog, N = 1, Score = 189, P = 6.5e-15

SWISSPROT:PLM_RAT PHOSPHOLEMMAN PRECURSOR., N = 1, Score = 185, P =
1.7e-14

>SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR.
Length = 92

HSPs:

Score = 196 (29.4 bits), Expect = 1.2e-15, P = 1.2e-15
Identities = 43/85 (50%), Positives = 56/85 (65%)

Query: 4 VLVLFLCSLLAPMVLASAAEKEKEMDPFHYDYQTLRIGGLVFAVVLFSVGILLILSRRCKC 63
+LVF LL +AE KE DPF YDYQ+L+IGGLV A +LF +GIL++LSRRC+C
Sbjct: 7 ILVFCVGLLT---MAKAESPKEHDPFTYDYQSLQIGGLVIAGILFILGILVLSRRRC 62

Query: 64 SFNQKPRA--PGDEEAQVENLITANAT 88
FNQ+ R P +EE + I +T
Sbjct: 63 KFNQQQTGEPEDEEGTFRSSIRRLST 89

Pedant information for DKFZphfbr2_82i17, frame 2

Report for DKFZphfbr2_82i17.2

[LENGTH]	95
[MW]	10542.37
[pI]	5.05
[HOMOL]	SWISSPROT:PLM_HUMAN PHOSPHOLEMMAN PRECURSOR. 3e-15
[BLOCKS]	BL01310

DKFZphfbr2_82i24

group: nucleic acid management

DKFZphfbr2_82i24 encodes a novel 547 amino acid protein with similarity to DEAD-box superfamily ATP-dependent helicases.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis.

The novel protein contains a DEAD-box an ATP/GTP-binding site motif A (P-loop, interacting with one of the phosphate groups of the nucleotide) and a leucine zipper. Mutations in the closely related *Drosophila* Hlc gene result in lethality in homozygotes. Therefore the new protein seems to be critical involved in RNA processing in eukaryotic cells.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to DEAD-box subfamily ATP-dependent helicase

complete cDNA, complete cds, EST hits
potential Start at Bp 9 matches Kozak consensus PyNnatgG,
[PFAM] Helicases conserved C-terminal domain
[PFAM] DEAD and DEAH box helicases

Sequenced by DKFZ

Locus: /map="720_A_3; 758_H_4; 772_E_3; 804_A_5; 175.5 cR from topFT of Chr7 linkage group"

Insert length: 1860 bp

Poly A stretch at pos. 1850, polyadenylation signal at pos. 1829

```
1 AGCAGCGCCA TGGAGGACTC TGAAGCACTG GGCTTCGAAC ACATGGGCCT
51 CGATCCCCGG CTCCTTCAGG CTGTCAACCGA TCTGGGCTGG TCGCGACCTA
101 CGCTGATCCA GGAGAAGGCC ATCCCACTGG CCCTAGAAGG GAAGGACCTC
151 CTGGCTCGGG CCCGCACGGG CTCCGGGAAG ACGGCCGCTT ATGCTATTCC
201 GATGCTGCAG CTGTTGCTCC ATAGGAAGGC GACAGGTCCG GTGGTAGAAC
251 AGGCAGTGAG AGGCCTTGTT CTTGTTCCTA CCAAGGAGCT GGCACGGCAA
301 GCACAGTCCA TGATTACGCA GCTGGCTACC TACTGTGCTC GGGATGTCCG
351 AGTGGCCAAT GTCTCAGCTG CTGAAGACTC AGTCTCTCAG AGAGCTGTGC
401 TGATGGAGAA GCCAGATGTG GTAGTAGGGA CCCCATCTCG CATATTAAGC
451 CACTTGCAGC AAGACAGCCT GAAACTTCGT GACTCCCTGG AGCTTTTGGT
501 GGTGGACGAA GCTGACCTTC TTTTTCCTT TGGCTTTGAA GAAGAGCTCA
551 AGAGTCTCCT CTGTCACTTG CCCCGGATTT ACCAGGCTTT TCTCATGTCA
601 GCTACTTTTA ACGAGGACGT ACAAGCACTC AAGGAGCTGA TATTACATAA
651 CCCGGTTACC CTTAAGTTAC AGGAGTCCCA GCTGCCTGGG CCAGACCACT
701 TACAGCAGTT TCAGGTGGTC TGTGAGACTG AGGAAGACAA ATTCTCCTCTG
751 CTGTATGCCC TGCTCAAGCT GTCATTGATT CGGGGCAAGT CTCTGCTCTT
801 TGTCAACACT CTAGAACGGA GTTACCGGCT ACGCCTGTTC TTGGAACAGT
851 TCAGCATCCC CACCTGTGTG CTCAATGGAG AGCTTCCACT GCGCTCCAGG
901 TGCCACATCA TCTCACAGTT CAACCAAGGC TTCTACGACT GTGTCAATAGC
951 AACTGATGCT GAAGTCTTGG GGGCCCCAGT CAAGGGCAAG CGTCGGGGCC
1001 GAGGGCCCAA AGGGGACAAG GCCTCTGATC CGGAAGCAGG TGTGGCCCCG
1051 GGCATAGACT TCCACCATGT GTCTGCTGTG CTCAACTTTG ATCTTCCCCC
1101 AACCCCTGAG GCCTACATCC ATCGAGCTGG CAGGACAGCA CGCGCTAACA
1151 ACCCAGGCAT AGTCTTAACC TTTGTGCTTC CCACGGAGCA GTTCCACTTA
1201 GGCAAGATTG AGGAGCTTCT CAGTGGAGAG AACAGGGGCC CCATTCTGCT
1251 CCCTTACCAG TTCCGGATGG AGGAGATCGA GGGCTTCCGC TATCGCTGCA
1301 GGGATGCCAT GCGCTCAGTG ACTAAGCAGG CCATTGCGGA GGCAAGATTG
1351 AAGGAGATCA AGGAAGAGCT TCTGCATTCT GAGAAGCTTA AGACATACTT
1401 TGAAGACAAC CCTAGGGACC TCCAGCTGCT GCGGCATGAC CTACCTTTGC
1451 ACCCCGCACT GGTGAAGCCC CACCTGGGCC ATGTTCTCTGA CTACCTGGTT
1501 CCTCTGCTC TCCGTGGCCT GGTACGCCCT CACAAGAAGC GGAAGAAGCT
1551 GTCTTCTCTT TGTAGGAAGG CCAAGAGAGC AAAGTCCCAG AACCCACTGC
1601 GCAGCTTCAA GCACAAAGGA AAGAAATTCA GACCCACAGC CAAGCCCTCC
1651 TGAGGTTGTT GGGCCTCTCT GGAGCTGAGC ACATTGTGGA GCACAGGCTT
1701 ACACCCCTTC TGGACAGGCG AGGCTCTGGT GCTTACTGCA CAGCCTGAAC
1751 AGACAGTTCT GGGGCCGCGA GTGCTGGGCC CTTAGCTCC TTGGCACTTC
1801 CAAGCTGGCA TCTTGCCCTT TGACAACAGA ATAAAAATTT TAGCTGCCCC
1851 AAAAAAAAAA
```

BLAST Results

Entry HSG05793 from database EMBL:

human STS WI-6581.

Length = 206

Minus Strand HSPs:

Score = 992 (148.8 bits), Expect = 6.0e-38, P = 6.0e-38

Identities = 204/208 (98%), Positives = 204/208 (98%), Strand = Minus /
P1

Entry AC004938 from database EMBL:

Homo sapiens clone DJ0971C03; HTGS phase 1, 18 unordered pieces.

Score = 1269, P = 6.5e-202, identities = 269/282

12 exons Bp -87920-93706 (matching 1-1497)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 10 bp to 1650 bp; peptide length: 547

Category: strong similarity to known protein

Classification: Nucleic acid management

Prosites motifs: ATP_GTP_A (51-59)

LEUCINE_ZIPPER (149-171)

```

1 MEDSEALGFE HMGLOPRLQ AVTDLGWSRP TLIQEKAIP L ALEGKDLLAR
51 ARTGSGKTAA YAIPMLQLL HRKATGPVVE QAVRGLVLVP TKELARQAQS
101 MIQQLATYCA RDVRVANVSA AEDSVSQRAV LMEKPDVVVG TPSRILSHLQ
151 QDSIKLRDSL ELLVVDEADL LFSFGFEEEL KSLLCPLPRI YQAFILMSATF
201 NEDVQALKEL ILHNPVTLKL QESQLPGPDQ LQQFQVVCET EEDKFLLLYA
251 LLKLSLRGK SLLFVNTLER SYRLRLFLEQ FSIPTCVLNG ELPLRSRCHI
301 ISQFNQGFYD CVIATDAEVL GAPVKGKRRG RGPKGDKASD PEAGVARGID
351 FHHVSAVLNF DLPPTPEAYI HRAGRTARAN NPGIVLTFVL PTEQFHLGKI
401 EELLSGENRG PILLPYQFRM EEIEGFRYRC RDAMRSVTKQ AIREARLKEI
451 KEELLHSEKL KTYFEDNPRD LQLLRHDLPL HPAVVKPHLG HVPDYLVPFA
501 LRGLVRPHKK RKKLSSSCRK AKRAKSNPL RSFKHKGKKF RPTAKPS

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfbr2_82i24, frame 1

TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), wacław (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds., N = 1, Score = 1230, P = 3.2e-125

TREMBL:SPCC1494_6 gene: "SPCC1494.06c"; product: "atp dependent helicase"; S.pombe chromosome II cosmid c1494., N = 2, Score = 753, P = 2.5e-113

PIR:S51412 hypothetical protein YLR276c - yeast (Saccharomyces cerevisiae), N = 2, Score = 711, P = 8.2e-117

TREMBL:AF025451_2 gene: "C24H12.4"; Caenorhabditis elegans cosmid C24H12., N = 2, Score = 564, P = 2.7e-99

>TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (iby), wacław (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst), and la costa (lcs) genes, complete cds.
Length = 560

HSPs:

Score = 1230 (184.5 bits), Expect = 3.2e-125, P = 3.2e-125

Identities = 251/497 (50%), Positives = 344/497 (69%)

Query: 9 FEHMGDPRLLQAVTDLGWSRPTLIQEKAIPALEGKDLLARARTGSGKTAAYAI PMLQL 68
F + LD R+L+AV LGW +PTLIQ AIPL LEGKD++ RARTGSGKTA YA+P++Q
Sbjct: 11 FHELELDQRILKAVAQLGWQQPTLIQSTAIPLLEGGKDVVRARTGSGKTATYALPLIQK 70

Query: 69 LLHRKATGPVVEQAVRGLVLPVKELARQAQSMIQQLATYCARDVRVANVS-AAEDSVSQ 127
+L+ K EQ V +VL PTKEL RQ++ +I+QL C + VRVA+++ ++ D+V+Q
Sbjct: 71 ILNSKLNAS--EQYVSAVVLAPTKELCRQSRKVEQLVESCGKVVVRADIADSSNDTVTQ 128

Query: 128 RAVLMEKPDVVVGTPSRILSHLQQDSLKLKRLDSLELLVVDEADLLFSFGFEEELKSLCHL 187
R L E PD+VV TP+ +L++ + S+ +E LVVDEADL+F++G+E++ K L+ HL
Sbjct: 129 RHALSESPDIVVATPANLLAYAEAGSVVDLKHVETLVVDEADLVFAYGYEKDFKRLIKHL 188

Query: 188 PRIYQAFILMSATFNEDVQALKELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKELL 247
P IYQA L+SAT +DV +K L L+NPVTLKL+E +L DQL +++ E E DK +
Sbjct: 189 PPIYQAVLVSATLTDVVVRMKGCLNPNVTLKLEPELVPPQDQLSHQRILAE-ENDKPAI 247

Query: 248 LYALLKLSLIRGKSLLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHIIISQFNQG 307
LYALLKL LIRGKS++FVN+++R Y++RLFLEQF I CVLN ELP R H ISQFN+G
Sbjct: 248 LYALLKRLIRGKSII FVNSTORCYKVRFLFLEQFGIRACVLNSEL PANIRIHTISQFNKG 307

Query: 308 FYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPE 367
YD +IA+D + P G + K ++ D E+ +RGIDF V+ V+NFD P
Sbjct: 308 TYDIIIASDEHHMEKP--GGKSATNRKSPRSGDMESSASRGIDFQCVNNVINFDFFPRDVT 365

Query: 368 AYIHRAGRTARANNPGIVLTFVLPTQFHLGKIEELL---SGENRGPIILPYQFRMEEI 423
+YIHRAGRTAR NN G VL+V E +E+ L + + I+ YQF+MEE+
Sbjct: 366 SYIHRAGRTARGNNGSVLSFVSMKESKVNDSEVKLCLDSFAAQEGEQIKNYQFKMEEV 425

Query: 424 EGFYRCDAMRSVTQKQAIAREARLKEIKEELLHSEKLTXYFEDNPRDLQLLRHDLPLHPA 483
E FRYR +D R+ T+ A+ + R++EIK E+L+ EKLK +FE+N RDLQ LRHD PL
Sbjct: 426 ESFRYRAQDCWRAATRVAVHDTREIKIEILNCEKLAFFEEENKRDQLALRHDKPLRAI 485

Query: 484 VVKPHLGHVPDYLVPPALRGLV 505
V+ HL +P+Y+VP AL+ +V
Sbjct: 486 KVQSHLSDMPEYIVPKALKRVV 507

Pedant information for DKFZphfbr2_82i24, frame 1

Report for DKFZphfbr2_82i24.1

[LENGTH] 547
[MW] 61589.88
[pI] 9.34
[HOMOL] TREMBL:AF017777_10 gene: "hlc"; product: "helicase"; Drosophila melanogaster
tweety (tty), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent
bystander (iby), waclaw (waw), bobby sox (bbx), sluggish (slg), helicase (hlc), misato (mst),
and la costa (lcs) genes, complete cds. 1e-121
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR276c] 1e-109
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 2e-42
[FUNCAT] 04.01.04 rRNA processing [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YLL008w] 8e-40
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YKR059w] 3e-39
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YKR059w] 3e-39
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 3e-35
[FUNCAT] 04.05.03 mRNA processing (splicing) [S. cerevisiae, YPL119c] 3e-29
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 4e-29
[FUNCAT] 1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 1e-27
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-27
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 4e-21
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 1e-05
[BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW] nucleus 4e-34
[PIRKW] RNA binding 7e-41
[PIRKW] DEAD box 2e-38
[PIRKW] transmembrane protein 9e-20
[PIRKW] DNA binding 8e-23
[PIRKW] ATP 1e-107
[PIRKW] purine nucleotide binding 2e-38
[PIRKW] P-loop 1e-107
[PIRKW] hydrolase 2e-35
[PIRKW] protein biosynthesis 2e-38
[PIRKW] ATP binding 7e-43

```

[SUPFAM]      WW repeat homology 1e-26
[SUPFAM]      DEAD/H box helicase homology 1e-107
[SUPFAM]      unassigned DEAD/H box helicases 1e-107
[SUPFAM]      ATP-dependent RNA helicase DBP1 3e-31
[SUPFAM]      ATP-dependent RNA helicase DHH1 2e-35
[SUPFAM]      translation initiation factor eIF-4A 2e-38
[SUPFAM]      tobacco ATP-dependent RNA helicase DB10 1e-26
[PROSITE]     ATP_GTP_A      1
[PROSITE]     LEUCINE_ZIPPER 1
[PFAM]        Helicases conserved C-terminal domain
[PFAM]        DEAD and DEAH box helicases
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY      9.87 %

```

```

SEQ      MEDSEALGFHEHGLDPRLLQAVTDLGWSRPTLIQEKAIPALALEGKDLLARARTGSGKTAA
SEG      .....
PRD      cccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ      YAIPMLQLLLHRKATGPVVEQAVRGLVLPVKELARQAQSMIQLATYCARDVVRVANVSA
SEG      .....
PRD      ehhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      AEDSVSQRAVLMKPDVVVGTPSRILSHLQDQSLKLRDSLELLVVDADLLFSFGFEEEL
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      KSLLLCHLPRIYQAFILMSATFNEDVQALKELILHNPVTLKLQESQLPGPDQLQQFQVVCET
SEG      .....
PRD      hhhhhhhccccchhhhhhhhhccccchhhhhhhhhhhccccccccccccccccchhhhhhhhhhh

SEQ      EEDKFLLLYALLKLSLIRGKSLFVNTLERSYRLRLFLEQFSIPTCVLNGELPLRSRCHI
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccchhhhhhhhh

SEQ      ISQFNQGFYDCVIATDAEVLGAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNF
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      DLPPTPEAYIHRAGRTARANNPGIVLTVLPTQFHLGKIEELLSGENRGPIILPYQFRM
SEG      .....
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ      EEIEGFRYRCRDAMRSVTKQAIPEARLKEIKEELLHSEKLKTYFEDNPRDLQLLRHDLPL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhcccc

SEQ      HPAVVKPHLGHPDYLVPFALRGLVRPHKKRKLSSSSCRKAKRAKSNPLRSFKHKGKKF
SEG      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      RPTAKPS
SEG      .....
PRD      ccccccc

```

Prosites for DKFZphfbr2_82i24.1

```

PS00017      51->59      ATP_GTP_A      PDOC00017
PS00029      149->171    LEUCINE_ZIPPER      PDOC00029

```

Pfam for DKFZphfbr2_82i24.1

```

HMM_NAME      DEAD and DEAH box helicases

HMM      *gLpPWILRnIyeMGFEkPTPIQQQAIPiLeGRDVMACAQTGSGKTAAAF
Query      13      GLDPRLLQAVTDLGWSRPTLIQEKAIPALALEGKDLLARARTGSGKTAAAY      61

HMM      1IPMLQHIDwdP...WpqpPQdPrALILAPTRELAMQIEEcRkFgkHmN
Query      62      AIPMLQLLLHRKATGPVVEQA-VRGLVLPVKELARQAQSMIQLATYCA      110

HMM      g.IRImciYGGtnMRdQMRmLeRGpPHIVATPGRLLIDHIERgtldLDr.
Query      111      RDVRVANVSAEDSVSQRAVLMKPDVVVGTPSRILSHLQDQSLKLRDS      159

HMM      IeMLVMDEADRLMDMGFIDQIRrIMrqIPmPwNRQTMMSATMPdeIqEL
Query      +E LV DEAD +++ GF++++ ++ +P + Q + SAT+ +++Q L

```

```

Query      160 LELLVVDEADLLFSFGFEEELKSLCHLP--RIYQAFILMSATFNEDVQAL  207
HMM          ARrFMRNPiRinIdMdElTtnEnikQwYiyVerEMWKfdCLcrLIe*
          + +++NP+ + + +++L + ++Q+ +++E E++KF +L+ L++
Query      208 KELILHNPVTLKLQESQLPGPDQLQQFQVVCETEEDKFLLLYALLK  253

HMM_NAME      Helicases conserved C-terminal domain
HMM          *EilleeWLknlGIrvmYIHGdMpQeERdeImddFNnGEynVLicTDV...
          +L+ +L++ I+++++ G +P + R I+ +FN+G Y++ I+TD+
Query      272 YRLRLFLEQFSIPTCVLNGELPLRSRCHIISQFNQGFYDCVIATDAEVL  320
HMM          .....ggrGIDIPdVNHVINYDMPWNPEqYI
          +RGID+ V+ V N+D+P +PE YI
Query      321 GAPVKGKRRGRGPKGDKASDPEAGVARGIDFHHVSAVLNFDLPPTPEAYI  370
HMM          QRIGRTgRIG*
          +R+GRT+R++
Query      371 HRAGRTARAN  380

```

DKFZphfbr2_82m16

group: brain derived

DKFZphfbr2_82m16 encodes a novel 289 amino acid protein with very weak similarity to A.thaliana F28A23.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of brain-specific genes.

similarity to A.thaliana F28A23.140

complete cDNA, complete cds, few EST hits
many ATGs in front of the ORF
TRANSMEMBRANE 1

Sequenced by DKFZ

Locus: /map="4"

Insert length: 2715 bp
Poly A stretch at pos. 2705, polyadenylation signal at pos. 2687

```
1 AGAGGAGGGG AGAGGACTGG GGAGCCGAGC CAGAGCCGGG CTGCCTGCCA
51 CCCGGGTGCT CGTCCGCTAG CTGGGGAGGA GCGCTCCACC CGCAACTGAC
101 AAAGGATGGG AGAATGCCCG CGCCCCGGGA TGCCGGCCCG ACGCAGCCTG
151 GCGGCCGCGCT GAGCTACTTC ACCCTCCGCC GGTAAGTGAC TGCAAACATC
201 ATTCATTCAA TCAGCCTCAC TGGGAGCCCC TTCTCTCCGG CTGGTAGTCC
251 TGGGCGGCTT GTCCCTGATC CCGAGCGGGG CTTGGCACAG CATCAGCCCT
301 GGAGGGCAGG CAGCAGGTGC CTTTGCCTGG TGGGTCCACT GGGGAGCGTG
351 GCTGGGGTTC GCGGCGGGTG CTGCCACCCA ACCTGCGGGC GCGGGGCTCG
401 CCCAGTAGGC GCCTCTCTGG TGAGAGGAGG CGGCTCCAGC CCGCATCCTG
451 GGGTAGTTGC TACTATTGGC CCCCAGCGCC CGCTCTGCGC GCGCGCGGTT
501 TCTGGCGGAT CCCCAGTGCG CGGCGCGCTG TTTACACCGG CGTGGTACTA
551 GTACACGGAG CGCACCCCTC GGAAAGCGCG GAGTCGATGA CAGCCACTTC
601 ACAGGCTCAC GCGCTCCTAG TGTGGGCTTG AAGGGGACGG GGACCGATTA
651 CCAAAGGAGA GCGCTGAGTA CGGAAGACAC AGGGCAGCCT TTGTCTTGGG
701 TTTAGCGCTG ATGCGCTCAA CCCTGAGTCG GGTTCAGTGC AACTGTTGTG
751 TCCGATTTCT GTTCCCTGCA ACCGCCCTCC TGGGCGAGAG ATGTCAATTG
801 GTTCTCTGCC CCAGCGGGAC TGAGAGCTGG GACTTAAGAC GCCAGGAGGG
851 TCCTGCGCTC ACGGGAATG TACCCCAAAA GAAGCTGAG AGAATATACT
901 CAAGTGTCTT GCTGTGATTA AACAAGACTG CTGTATTTTA ATTTAGAAAA
951 TTGAAAAGGG ATAGGAGGAA GGGGAAAATG CTGGGCTGGT GTGAAGCGAT
1001 AGCCCGTAAC CCTCACAGAA TTCCAAACAA CACGCGAACA CCCGAGATCT
1051 CAGGGGATTT GGCTGACGCC TCACAAACCT CCACATTGAA TGAATAATCT
1101 CCAGGGCGAT CTGCAAGTCG ATCAAGTAAC ATTTCAAAAG CAAGCAGCCC
1151 AACAACAGGG ACAGCTCCCA GGAGCCAGTC AAGGTTGTCT GTCTGTCCAT
1201 CCACCTCAGG CATCTGCAGA ATCTGTCACT GCGAAGGGGA TGAAGAGAGC
1251 CCCCTCATCA CACCCTGTGC CTGCACTGGG AACTGCGCT TTGTCCACCA
1301 GTCTTGCCTC CACCAAGTGA TAAAGAGCTC AGATACACGC TGCTGTGAGC
1351 TCTGCAAGTA TGACTTCATA ATGGAGACCA AGCTCAAAAC CCTCGGAAG
1401 TGGGAGAAAC TACAGATGAC CACAAGTGAA AGGAGGAAAA TATTCTGCTC
1451 TGTACACATC CACGTAATCG CGATCACCTG TGTGGTTTGG TCTTTGTATG
1501 TATTGATAGA CCGGACAGCG GAGGAAATCA AGCAAGGCAA TGACAATGGT
1551 GTCCTTGAAT GGCCATTTTG GACAAAACCT GTTGTGGTAG CCATTGGCTT
1601 CACAGGAGGT CTTGTCTTCA TGTACGTACA GTGTAAGATC TATGTTCACT
1651 TGTGGCGCAG GCTGAAGGCC TACAACCGTG TGATCTTTGT AAAAAATTGC
1701 CCAGACACTG CCAAAAAACT GGAGAAGAAC TTCTCATGTA ATGTAAACAC
1751 AGACATCAAA GATGCTGTGG TAGTGCTGTG ACCACAAACA GGTGCAAAAT
1801 CACTGCCATC TGCAGAGGGT GGCCCCCCTG AAGTTGTATC AGTCTGATGG
1851 AACCTGTTGG GAGTTTCTTC ACCGAAGAAT ATCTTTCTAG CCCTCAGCCA
1901 CTACAAATGA CAGAAGTGAC CTTGAATTAT TTAATCCCTT CAGCTCCTCC
1951 TTTCTCCTAC TGACACATTT TTCCTGACTT TGTTCAAAGA GGAAGGAGA
2001 AAAACAAACA AACAGACCAA ATGCCAGGA GCCCATGAAG TAATAGCGTA
2051 AAGTAAAGTA TGATATGGAA ATGTGAAGTT TGCAAGAGAA TGATTTCAC
2101 GACAATTTAG AACTACTGGG GCAATGAATG CTTTATAGGA GTAATCAAAG
2151 ATTAATATGA CCCATGATAC TCTTCTTCA AGTAACAGGG GAAAGATTCA
2201 AGAATACAGA CTTGAATTGC GATGTGTATT ACTTCTAGGG CCTTGTAAATG
2251 TTAAGTGTCT CATCTGGAAA TAATAACTAA CATATTGGT TTTAAGCCTG
2301 AAATTGTCTG CATTATCCCT AAGTCACATT GGAAGTGAAC TTGGAGGATG
2351 CATATTTTGA TATGCTTTGA CAGCTAACAG ATTTGTATGG TTTAGTGGAG
2401 TCTGGTTATT TTGACAGATG CATGTTTTTT TTAATATAGT GCAATATACA
2451 TTTGAAGACA TTGATATTTG GAATTAATTA TGTGTGTTTA AGTCACGCAA
2501 AAGATTTTCA GAAAATGTTT GGATATAATT AGCTCTGTTA AATACCCACA
2551 GAAGTGTAT CAGGTCTTAT ATTTATTTTC ATCTGGTTCC TCTAATACAG
```

2601 TGCTGTCCAA TAGAAACACA ACAGCCACAA ATGCAGGCCA CAGATGCAAA
 2651 TATTTAAGT CCCAGTAGCC CTATTTTAAA AAGTAAAAAT AAATGTTTGT
 2701 TTGTTAAAAA AAAAA

BLAST Results

Entry G37457 from database EMBLNEW:
 SHGC-57357 Human Homo sapiens STS genomic.
 Length = 458
 Plus Strand HSPs:
 Score = 2116 (317.5 bits), Expect = 4.3e-91, P = 4.3e-91
 Identities = 444/456 (97%)

Medline entries

No Medline entry

Peptide information for frame 3

1 MLGWCEAIAR NPHRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS
 51 NISKASSPTT GTAPRSQSRL SVCSTQDIC RICHCEGDEE SPLITPCRCT
 101 GTLRFVHQSC LHWIKSSDT RCCELCKYDF IMETKLKPLR KWEKLQMTTS
 151 ERRKIFCSVT FHVIAITCVV WSLYVLIDRT AEEIKQGNND GVLEWPFWTK
 201 LVVVAIGFTG GLVFMVQCK VYVQLWRRLK AYNRVIFVQN CPDTAKKLEK
 251 NFSCNVNTDI KDAVVVPVPQ TGANSLPSAE GGPPEVVS

ORF from 978 bp to 1844 bp; peptide length: 289
 Category: similarity to unknown protein

BLASTP hits

Entry AB011169_1 from database TREMBL:
 gene: "KIAA0597"; product: "KIAA0597 protein"; Homo sapiens mRNA for
 KIAA0597 protein, partial cds.
 Score = 188, P = 6.0e-12, identities = 30/54, positives = 38/54

Entry SPBC14F5_7 from database TREMBL:
 gene: "SPBC14F5.07"; product: "hypothetical protein"; S.pombe
 chromosome II cosmid c14F5.
 Score = 185, P = 1.9e-11, identities = 29/53, positives = 38/53

Entry CEY57A10B_1 from database TREMBL:
 gene: "Y57A10B.1"; Caenorhabditis elegans cosmid Y57A10B
 Score = 171, P = 2.6e-10, identities = 40/107, positives = 58/107

Alert BLASTP hits for DKFZphfbr2_82ml6, frame 3

TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII
 project), N = 1, Score = 198, P = 3.4e-13

>TREMBL:ATF28A23_14 gene: "F28A23.140"; product: "putative protein";
 Arabidopsis thaliana DNA chromosome 4, BAC clone F28A23 (ESSAII project)
 Length = 1,051

HSPs:

Score = 198 (29.7 bits), Expect = 3.4e-13, P = 3.4e-13
 Identities = 38/103 (36%), Positives = 61/103 (59%)

Query: 28 LADASQTSTLNEKSPGRSASRS-SNISKASSPTTGTAPRSQSRLSVCSTQDICRICHCE 86
 +++ S +S+ + SP +++ SN+ A S TG+ +D+CRIC
 Sbjct: 20 VSEPSVSSSSSSSPNQASPNPFSNMDPAVSTATGSGRYVDDDE-----DEEDVCRICRNP 74
 Query: 87 GDEESPLITPCRCTGTLRFVHQSC LHWIKSSDTRCCCELCKYDF 130
 GD ++PL PC C+G+++FVHQ CL QW+ S+ R CE+CK+ F
 Sbjct: 75 GDADNPLRYPACSGSIKFFVHQDCLLQWLNHSNARQCEVCKHPF 118

Pedant information for DKFZphfbr2_82m16, frame 3

Report for DKFZphfbr2_82m16.3

```

[LENGTH]      289
[MW]           32308.36
[pI]           8.76
[HOMOL]        PIR:T00268 hypothetical protein KIAA0597 - human (fragment) 9e-14
[FUNCAT]        04.99 other transcription activities [S. cerevisiae, YIL030c] 4e-09
[PIRKW]         transmembrane protein 9e-08
[PROSITE]       MYRISTYL 1
[PROSITE]       CK2_PHOSPHO_SITE 4
[PROSITE]       TYR_PHOSPHO_SITE 1
[PROSITE]       PKC_PHOSPHO_SITE 3
[PROSITE]       ASN_GLYCOSYLATION 3
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY 6.57 %

```

```

SEQ  MLGWCEAIARNPHRIPNNTRTPETISGDLADASQTSTLNEKSPGRSASRSSNISKASSPTT
SEG  .....xxxxxxxxxxxxxxxxxxxxx...
PRD  cccchhhhhccccccccccccccccchhhhhhhcccccccccccccccccccccccccc

SEQ  GTAPRSQSRLSVCPTQDICRICHEGDEESPLITPCRCTGTLRFVHQSLHQQWIKSSDT
SEG  .....
PRD  cccccccccccccccccceeeeeccccccccccccccccccccceeeehhhhhhhhhcccc

SEQ  RCCELCKYDFIMETKLKPLRKWEKLQMTTSERRKIFCSVTFHVIAITCVVWSLYVLIDRT
SEG  .....
PRD  ceeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

SEQ  AEEIKQGNONGVLEWPFWTKLVVVAIGFTGGLVFMVYQCKVYVQLWRRLKAYNRVIFVQN
SEG  .....
PRD  cccccccccceeeehhhhheeeeeccccccccceeeehhhhhhhhhhhhhhhheeeee

SEQ  CPDTAKKLEKNFSCNVNTDIKDAVVVPVPTGANSLSAEGGPPEVVS
SEG  .....
PRD  ccchhhhhccccccccccccceeeeecccccccccccccccccccccc

```

Prosite for DKFZphfbr2_82m16.3

PS00001	17->21	ASN_GLYCOSYLATION	PDOC00001
PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	244->247	PKC_PHOSPHO_SITE	PDOC00005
PS00006	36->40	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00007	121->129	TYR_PHOSPHO_SITE	PDOC00007
PS00008	187->193	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfbr2_82m16.3)

DKF2phfbr2_82m6

group: signal transduction

DKF2phfbr2_82m6.3 encodes a novel 654 amino acid protein with similarity to murine sphingosine kinase.

Sphingosine kinase is a new type of lipid kinase, which is regulated by growth factors. The enzyme phosphorylates sphingosine, which subsequently exerts intracellular and extracellular actions. Intracellularly, sphingosine 1-phosphate (SPP) promotes proliferation and inhibits apoptosis. In yeast, survival of cells exposed to heat shock indicates is dependend on SPP. Extracellularly, SPP inhibits cell motility and influences cell morphology, effects that appear to be mediated by the G protein-coupled receptor EDG1.

The new protein can find application in modulating/blocking the shingosine kinase intracellular signal transmission pathway.

strong similarity to mouse "sphingosine kinase"

complete cDNA, complete cds, EST hits,
YLR260w/YOR171c Lcb5p/Lcb4p = long chain base kinases,
involved in biosynthesis of sphingolipids

Sequenced by DKF2

Locus: unknown

Insert length: 2875 bp

Poly A stretch at pos. 2865, polyadenylation signal at pos. 2838

```
1 AGTGTGGAG GTGAGGAGGC GGGGCTGGCA GGGCTAGTCG GGGCATCTGG
51 AAATTTCCGA CCCACGCTT CGGGCGTTTC CTTATCAGGT TCACCGCTCC
101 CTGATCTCGC GCTGCACTTC GTAGGCGCAG CCGCTGCTTG GGAAGTCCTA
151 CTTAAGAGCT GAAGGTCAGG CCAGGACAGT GAGACCTGAC TCCTTGCTCC
201 TACCAGCCTA CTATGGCTTA AGACCCAGGG CCAGGGTCCC GTTGATGTAA
251 CAGAGCAGAG GACCAGCAGA TGAATGGACA CCTTGAAGCA GAGGAGCAGC
301 AGGACCCAGG GCCAGACCAG GAGCTGACCG GGAGCTGGGG CCACGGGCGT
351 AGGAGCACCC TGGTCAGGGC TAAGGCCATG GCCCCGCCCC CACCGCCACT
401 GGCTGCCAGC ACCTCGCTCC TCCATGGCGA GTTTGGCTCC TACCCAGCCC
451 GAGGCCACAG CTTTGCCCTC ACCCTTACAT CGCAGGCCCT GCACATACAG
501 CGGCTGGGCC CCAAACCTGA AGCCAGGCCC CGGGGTGGCC TGGTCCCCTT
551 GCGCGAGGTC TCAGGCTGCT GCACCTSCG AAGCCGCAGC CCCTCAGACT
601 CAGCGGCTTA CTTCTGCATC TACACCTACC CTCGGGGCCG GCGCGGGGCC
651 CGGCGCAGAG CCACTCGCAC CTTCGGGCGA GATGGGGCCG CCACCTACGA
701 AGAGAACCGT GCCGAGGCCC AGCGCTGGGC CACTGCCCTC ACCTGTCTGC
751 TCCGAGGACT GCCACTGCCC GGGGATGGGG AGATCACCCC TGACCTGCTA
801 CCTCGGCCCG CCGGTTGCT TCTATTGGTC AATCCCTTTG GGGGTGCGGG
851 CTTGGCTGCG CAGTGGTGTA AGAACCCAGT GCTTCCCATG ATCTCTGAAG
901 CTGGGCTGTC CTTCAACCTC ATCCAGACAG AACGACAGAA CCACGCCCGG
951 GAGCTGGTCC AGGGGCTGAG CTTGAGTGAG TGGGATGGCA TCGTCACGGT
1001 CTCGGGAGAC GGGCTGCTCC ATGAGGTGCT GAACGGGCTC CTAGATCGCC
1051 CTGACTGGGA GGAAGCTGTG AAGATGCCTG TGGGCATCCT CCCCTGCGGC
1101 TCGGGCAACG CGCTGGCCCG AGCAGTGAAC CAGCACGGGG GATTTGAGCC
1151 AGCCCTGGGC CTCGACCTGT TGCTCAACTG CTCACTGTTG CTGTGCCGGG
1201 GTGGTGGCCA CCCACTGGAC CTGCTCTCCG TGACGCTGGC CTCGGGCTCC
1251 CGCTGTTTCT CTTTCTGTG TGTGGCCTGG GGCTTCGTGT CAGATGTGGA
1301 TATCCAGAGC GAGCGCTTCA GGGCCTTGGG CAGTGCCCGC TTCACACTGG
1351 GCACGGTGCT GGGCCTCGCC ACACCTGCAC CTTACCGCGG ACGCCTCTCC
1401 TACCTCCCGC CCACTGTGGA ACCTGCCTCG CCCACCCCTG CCCATAGCCT
1451 GCCTCGTGCC AAGTCGGAGC TGACCCTAAC CCCAGACCCA GCCCCGCCCA
1501 TGGCCCACTC ACCCCTGCAT CGTTCTGTGT CTGACCTGCC TCTTCCCTCG
1551 CCCAGCCTG CCCTGGCCTC TCCTGGCTCG CCAGAACCCC TGCCCATCCT
1601 GTCCCTCAAC GGTGGGGGCC CAGAGCTGGC TGGGGACTGG GGTGGGGCTG
1651 GGGATGCTCC GCTGTCCCCG GACCCACTGC TGTCTTCACC TCCTGGCTCT
1701 CCCAAGGCAG CTCTACACTC ACCCGTCTCC GAAGGGGCCC CCGTAATTCC
1751 CCCATCTCTT GGGCTCCAC TTCCACCCC TGATGCCCGG GTAGGGGCCT
1801 CCACCTGCGG CCCGCCGAC CACTGCTGTC CTCCGCTAGG CACCCGCTG
1851 CCCCAGACT GGGTGACGCT GGAGGGGGAC TTTGTGCTCA TGTGGCCAT
1901 CTGCGCCAGC CACCTAGGCG CTGACCTGGT GGCAGCTCCG CATGCGCGCT
1951 TCGACGACGG CTTGGTGAC CTGTGCTGGG TGCGTAGCGG CATCTCGCGG
2001 GCTGCGCTGC TGGCCTTTT CTTGGCCATG GAGCGTGGTA GCCATTCTAG
2051 CCTGGGCTGT CCGCAGCTGG GCTACGCCCG GGCCCGTGCC TTCCGCCTAG
2101 AGCCGCTCAC ACCACGCGGC GTGCTCACAG TGGACGGGGA GCAGGTGGAG
2151 TATGGGCGCG TACAGGCACA GATGCACCTT GGCATCGGTA CACTGCTCAC
2201 TGCGCCTCCT GGCTGCCCGG GCGGGGAGCC CTGAAACTAA ACAAGCTTGG
2251 TACCCGCCCG GGGCGGGGCC TACATTCCAA TGGGGCGGAG CCTGAGCTAG
2301 GGGGTGTGGC CTGGCTGTCA GAGTGTGTGT GGCAGGGGCC CTGGCCCCGT
```

```

2351 CTCAGGATTG CGCTCGCTTT CATGGGACCA GACGTGATGC TGAAGGTGG
2401 GCGTCGTAC GGTAAAGAG AAATGGGCTC GTCCCGAGGG TAGTGCCTGA
2451 TCAATGAGGG CGGGGCTGG CGTCTGATCT GGGCCGCC TTACGGGGCA
2501 GGGCTCAGTC CTGACGCTTG CCACCTGCTC CTACCCGCC AGGATGGCTG
2551 AGGGCGGAGT CTATTTTACG CGTCGCCCAA TGACAGGACC TGAATGTAC
2601 TGGCTGGGGT AGGCCTCAGT GAGTCGGCCG GTCAGGGCCC GCAGCCTCGC
2651 CCCATCCACT CCGGTGCCTC CATTTAGCTG GCCAATCAGC CCAGGAGGGG
2701 CAGGTTCCCC GGGGCCGGCG CTAGGATTG CACTAATGTT CCTCTCCCG
2751 CGGGTGGGGG CGGGGAAATT CATATCCCCT GTTCGTCTCA TGCCTCTCCT
2801 CCGTCCCAA TCTAAAAGC AATTGAAAAG GTCTATGCAA TAAAGGCAGT
2851 CGCTTCATTC CTCTCAAAA AAAAA

```

BLAST Results

No BLAST result

Medline entries

99045661:

Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway.

98395082:

Molecular cloning and functional characterization of murine sphingosine kinase.

98241633:

Purification and characterization of rat kidney sphingosine kinase.

99178622:

Sphingosine 1-phosphate: a prototype of a new class of second messengers.

Peptide information for frame 3

```

1 MNGHLEAEEQ QQRPDQELT GSWGHGPRST LVRKAMAPP PPPLAASTSL
51 LHGEFGSYPA RGPREFALTLT SQALHIQRLR PKPEARPRGG LVPLAEVSGC
101 CTLRSRSPSD SAAYFCIYTY PRGRRGARRR ATRTFRADGA ATYEENRAEA
151 QRWATALTCL LRGLPLPGDG EITPDLLPRP PRLLLLVNPF GGRGLAWQWC
201 KNHVLPIMISE AGLSFNLIQT ERQNHARELV QGLSLSEWDG ITVSGDGLL
251 HEVLNGLDDR PDWEEAVKMP VGILPCGSGN ALAGAVNQHG GFEPALGLDL
301 LLNCSLLLCR GGGHPLDLLS VTLASGSRCF SFLSVANGFV SDVDIQSERF
351 RALGSARFTL GTVLGLATLH TYRGRLSYLP ATVEPASPTP AHSPLPRAKSE
401 LTLTPDPAPP MAHSPLHRSV SDLPPLPQP ALASPGSPEP LPILSLNGGG
451 PELAGDWGGA GDAPLSPDPL LSSPFGSPKA ALHSPVSEGA PVIPPSSGLP
501 LPTPDARVGA STCGPPDHLL PPLGTPLPPD WVTLEGDFVL MLAISPSHLG
551 ADLVAAPHAR FDDGLVHLCW VRSGISRAAL LRLFLAMERG SHFSLGCPQL
601 GYAAARAARFRL EPLTPRGVLT VDGEQVEYGP LQQMHPGIG TLLTGPPGCP
651 GREP

```

ORF from 270 bp to 2231 bp; peptide length: 654

Category: similarity to known protein

BLASTP hits

Entry SPAC4A8.7 from database TREMBL:

gene: "SPAC4A8.07c"; product: "hypothetical protein"; S.pombe chromosome I cosmid c4A8.

Score = 301, P = 7.9e-32, identities = 68/190, positives = 109/190

Entry CEC34C6_3 from database TREMBLNEW:

product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

>TREMBL:CEC34C6_3 product: "C34C6.5"; Caenorhabditis elegans cosmid C34C6

Score = 273, P = 9.0e-29, identities = 78/265, positives = 142/265

Entry S67059 from database PIR:

hypothetical protein YOR171c - yeast (Saccharomyces cerevisiae)

>TREMBL:SC55021.9 gene: "O3615"; product: "O3615p"; Saccharomyces cerevisiae cosmid pUOA1258 from chromosome 15R. >TREMBL:SCYOR170W_2

S.cerevisiae chromosome XV reading frame ORF YOR170w

Score = 253, P = 2.0e-25, identities = 70/234, positives = 116/234

Entry S51398 from database PIR:

hypothetical protein YLR260w - yeast (*Saccharomyces cerevisiae*)
>TREMBL:SCL8479_4 gene: "YLR260W"; product: "Ylr260wp"; *Saccharomyces cerevisiae* chromosome XII cosmid 8479.

Score = 251, P = 1.0e-24, identities = 62/198, positives = 103/198

Alert BLASTP hits for DKFZphfbr2_82m6, frame 3

TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1b) mRNA, complete cds., N = 2, Score = 615, P = 1.2e-92

TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds., N = 2, Score = 616, P = 2e-92

TREMBL:ATF18E5_16 gene: "F18E5.160"; product: "putative protein"; *Arabidopsis thaliana* DNA chromosome 4, BAC clone F18E5 (ESSAII project), N = 2, Score = 370, P = 6.8e-33

>TREMBL:AF068748_1 gene: "SPHK1a"; product: "sphingosine kinase"; *Mus musculus* sphingosine kinase (SPHK1a) mRNA, partial cds.
Length = 504

HSPs:

Score = 616 (92.4 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
Identities = 128/260 (49%), Positives = 173/260 (66%)

Query: 154 ATALTCLLRGLPLPGDGEITPDLLPRPPRLLLVNPFGGRLAWQWCKNHVLP MISEAGL 213
A C L + E LLPRP R+L+NP GG+G A Q ++ V P + EA +
Sbjct: 110 APVAPCQREPRDLAMEPECPRGLLPRPCRVLVLLNPQGGKGAQLQFQSRVQPFLEEAEI 169

Query: 214 SFNLIQTERQNHARELVQGLSLSEWDGIVTVSGDGLLHEVLNGLLDRPDWEEAVKMPVGI 273
+F LI TER+NHARELV L WD + +SGDGL+HEV+NGL++RPDWE A++ P+
Sbjct: 170 TFKLLITERKNHARELVCAEELGHWDAVMSGDGLMHEVNVGLMERPDWETAIQKPLCS 229

Query: 274 LPCGSGNALAGAVNQHGGEFEPALGLDLLNCSLLLCRGGGHPDLLSVTLASGSRCSFSL 333
LP GSGNALA +VN + G+E DLL+NC+LLLCR P++LLS+ ASG R +S L
Sbjct: 230 LPGGSGNALAASVNHYAGYEQVTNEDLLINCTLLCRRRLSPMNLSSLHTASGLRLYSVL 289

Query: 334 SVAWGFVSDVDIQSERFRALGSARFTLGTVLGLATLHTYRGRLSYLPATVEPASPTPAH 392
S++WGFV+DVD++SE++R LG RFT+GT LA+L Y+G+L+YLP TV AS PA
Sbjct: 290 SLSWGFVADVDESEKYRRLGEIRFTVGTFFRLASLRIYQGQLAYLPVGTV--ASKRPAS 347

Query: 393 SL-PRAKSELTLTPDPAPPMH 413
+L + + L P P +H
Sbjct: 348 TLVQKGPVDTHLVPLEEPVPSH 369

Score = 324 (48.6 bits), Expect = 2.0e-92, Sum P(2) = 2.0e-92
Identities = 72/160 (45%), Positives = 100/160 (62%)

Query: 499 LPLPTDARVGASTC---GPPDHLLPPLGTPLPPDWVTL-EGDFVLM LAISPSHLGADLV 554
LP+ T ++ AST GP D L PL P+P W + E DF+L+L + +HL ++L
Sbjct: 335 LPVGTVASKRPASTLVQKGPVDTHLVPLEEPVPSHWTVVPEQDFLLVLVLLHLSSELF 394

Query: 555 AAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSFSLGCPOLGYAAARAFLRLEPLT 614
AAP R + G++HL +VR+G+RAALLRLFLAM++G H L CP L + AFRLEP +
Sbjct: 395 AAPMGRCEAGVMHLFYVRAGVSRALLRLFLAMQKGKHMELDCPYLVHVPVVAFRLEPRS 454

Query: 615 PRGVLTVDGEQVEYGPLQAMHPGIGTLLTGPPGCP-GRE 653
RGV +VDGE + +Q Q+HP ++ G P GR+
Sbjct: 455 QRGVFSVDGELMVCEAVQGQVHPNYLWMVCGSRDAPSGRD 494

Score = 37 (5.6 bits), Expect = 3.6e-62, Sum P(2) = 3.6e-62
Identities = 8/20 (40%), Positives = 9/20 (45%)

Query: 459 GAGDAPLSPDPLSSPPGSP 478
G+ DAP D PP P
Sbjct: 485 GSRDAPSGRDSRRGPPPEEP 504

Pedant information for DKFZphfbr2_82m6, frame 3

Report for DKFZphfbr2_82m6.3

[LENGTH] 654
 [MW] 69207.45
 [pI] 6.47
 [HOMOL] TREMBL:AF068749_1 gene: "SPHK1b"; product: "sphingosine kinase"; Mus musculus
 sphingosine kinase (SPHK1b) mRNA, complete cds. 2e-50
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YLR260w]
 4e-20
 [PROSITE] AMIDATION 1
 [PROSITE] CAMP_PHOSPHO_SITE 1
 [PROSITE] MYRISTYL 12
 [PROSITE] CK2_PHOSPHO_SITE 6
 [PROSITE] TYR_PHOSPHO_SITE 1
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 1
 [KW] Alpha_Beta
 [KW] LOW_COMPLEXITY 20.18 %

SEQ MNGHLEAEEQQDQRPDQELTGSWGHGPRSTLVRKAMAPPPPLAASTSLHGEFGSYPA
 SEGXXXXXXXXXXXXX.....
 PRD cccchhhhhhhccccceccccceehhhhhcccccecccccecccccccccc

SEQ RGRPFALTLSQALHIQRLRPKEARPRGGLVPLAEVSGCCTLRSRSPSDAAYFCIYTY
 SEG
 PRD cccceehhhhhhhhhhhcccccccccecccccecccccecccccecccccc

SEQ PRGRRGARRRATRTFRADGAATYEENRAEAQRWATALTCLLRGLPLPGDGEITPDLLPRP
 SEGXXXXX.....
 PRD cccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhcccccccccccccccccc

SEQ PRLLLLVNPFGGRLAWQCKNHVLPMISEAGLSFNLIQTERQNHARELVQGLSLSEWDG
 SEGXXXXX.....
 PRD ceeeeeeccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccce

SEQ IVTVSGDGLLHEVLNGLLDRPDWEEAVKMPVGILPCGSGNALAGAVNQHGGEFALGLDL
 SEGXXXXX.....
 PRD eeeccccccccccccccccccccchhhhhccccceccccccccccccccccccccchhhhh

SEQ LLNCSLLLCRGGHPLDLLSVTLASGSRCSFSLVAVGFVSDVDIQSERFRALGSARFTL
 SEGXXXXXXXXXXXXX.....
 PRD hhhhhccccccccccccceccccceccccceccccceehhhhhhhhhhhhhhhhhc

SEQ GTVLGLATLHTYRGLSYLPATVEPASPTPAHSLPRAKSELTLTPDPAPMAHSPPLHRSV
 SEG
 PRD hhhhhhhhhhhcc

SEQ SDLPLPLQPALASPGSPEPLPILSLNGGPELAGDWGAGDAPLSPDPLSSPPGSPKA
 SEGXXXXXXXXXXXXX.....
 PRD cccccccccccccccccccccceccccccccccccccccccccccccccccccccccce

SEQ ALHSPVSEGAQVIPPSSGLPLTPDARVGASTCGPPDHLLPPLGTPLPDPDWTLEGDFVL
 SEGXXXXXXXXXXXXX.....
 PRD eccce

SEQ MLAISPSHLGADLVAAPHARFDDGLVHLCWVRSGISRAALLRLFLAMERGSFSLGCPQL
 SEG
 PRD eeeeeccccccccccccccccccccceccccchhhhhhhhhhhhhhhhhccccceccccch

SEQ GYAARAFRLEPLTPRGVLTVDGEQVEYGPLQAMHPGIGTLLTGPPGCPGREGP
 SEGXXXXXXXXXXXXX.....
 PRD hhhhhhhhhcccccccccecccccecccccccccccccccccecccccccccc

Prosites for DKF2phbr2_82m6.3

PS00001	303->307	ASN_GLYCOSYLATION	PDOC00001
PS00002	245->249	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	129->133	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	102->105	PKC_PHOSPHO_SITE	PDOC00005
PS00005	134->137	PKC_PHOSPHO_SITE	PDOC00005
PS00005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS00005	347->350	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	371->374	PKC_PHOSPHO_SITE	PDOC00005
PS00005	477->480	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006

PS00006	142->146	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00006	341->345	CK2_PHOSPHO_SITE	PDOC00006
PS00006	419->423	CK2_PHOSPHO_SITE	PDOC00006
PS00007	106->115	TYR_PHOSPHO_SITE	PDOC00007
PS00008	56->62	MYRISTYL	PDOC00008
PS00008	212->218	MYRISTYL	PDOC00008
PS00008	232->238	MYRISTYL	PDOC00008
PS00008	272->278	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	279->285	MYRISTYL	PDOC00008
PS00008	361->367	MYRISTYL	PDOC00008
PS00008	476->482	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	574->580	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	640->646	MYRISTYL	PDOC00008
PS00009	122->126	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfbr2_82m6.3)

DKFZphfkd2_1j9

group: kidney derived

DKFZphfkd2_1j9.3 encodes a novel 105 amino acid protein with high similarity to *Xenopus laevis* XLCL2 protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes.

strong similarity to XLCL2 protein, African clawed frog

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 2955 bp

Poly A stretch at pos. 2935, polyadenylation signal at pos. 2915

```
1 GGGGGGGGCT GAGTGCTCAG TGGAGAGCGG GGAGTTGTGT CCACCTTGCC
51 GACGTCGCTA GCCGTGGGGC TGTCCTGGGA AGGCGGACGG CGAGCGCCCG
101 GTGTCCGCAC TCGGCCGCCT GCCGTGCCCG TCTGCCCGCG TGTCATCCTC
151 ACTCGGGACG CAGGGACCGT TTTTAAATCA CAGGGGCGTG TGTCAGCCTG
201 CCCTAGGACT TCATGTCTAT ATATTTCCCC ATTCACTGCC CCGACTATCT
251 GAGATCGGCC AAGATGACTG AGGTGATGAT GAACACCCAG CCCATGGAGG
301 AGATCGGCCT CAGCCCCCGC AAGGATGGCC TTTCCTACCA GATCTTCCCA
351 GACCCGTCAG ATTTTGACCG CCGCTGCAAA CTGAAGGACC GTCTGCCCTC
401 CATAGTGGTG GAACCCACAG AAGGGGAGGT GGAGAGCGGG GAGCTCCGGT
451 GGGCCCTCGA GGAGTTCCCT GTCCAGGAGG ATGAGCAAGA TAACTGCGAA
501 GAGACAGCGA AAGAAAATAA AGAGCAGTAG AGTCCCTGTG GACTCCCATG
551 GGTATACCA GCCAGCATCT GTTCTGAAC TGTGTTTTTC CCATCATGAC
601 GGAAGAAGAG AGTGAGCCGC AATGTGTTCTG AAAATGTCAA ACGAGGCTTC
651 TGTGTTGCAC CTGCAGATCA CCGAGTTGGT TTTCTTTTCT TTTCTTGCTT
701 TTTTTTTTTT TTTGAAATTT GCCGAGCAGT GGAGCCCTCT GACAATTTGC
751 AAGGCCCTCT GAGAAAGGAA GCTGCTTAGA GCCAGGGGGT TAGTGGGTGA
801 GGGGAGCGAG TGCTGTTTTT GAGATCATTG TCTGAACCA GGCAGCCTAG
851 TAGAGGCAGT GGTGGGATTC CAATGGGTCT TGGTGGGTGG GAGGTGGGGC
901 ATGTGCAAAAG CAAGCAAGGA ACATTTGGGG TAAGAAAACA AACATGAGGC
951 AAAGAAAAAA ATACATGTTT TTAAGAAAAC ATTGAGCAGA GAACTGCGAGC
1001 CAGGATGCGC TCAGCAGACA TTCACCTCTG CCGCTGGGAC ATCAGAAAAC
1051 AAGTCTCTCA TCTCTCTCTC CAGTTTCAAC CACCCCAACC TTTGCTTTCA
1101 TTTCAAGTGT GTTGGTCTAT ATGACAGGGA GGAGAGTAAA GGAGAGCAGG
1151 AGCAATTGGC TGCCCTGCAAA GCCAGCTGGA GGTGAAGTGC AGGAAAGGAA
1201 AGGTACCCCC ATTCTACTCC ATGGCCTCTC TGCTCCACAG TGTGGTAGGC
1251 TCACATAGCC AGTGTGATCG GTTTTAAAGA GGCAGTGCTT TTCAGCTTTT
1301 CTCCCTGATA TATCCATTTT GCTTCCCAGC ACTTTTATAG AGTAGTGAGA
1351 GCACCTTCCTG CCCTTGTTGG AAGCCCCAGG GTGGACACTC AGCAGCAAGG
1401 TCTCTCCCTT AACTGCTGCC CTTCGAAGAC TTGCTCCCGA GATGGAGTGG
1451 GCGTGGTCTT CCAGGCTGGC CCTTCCTTCT CCTCACCGCC ACCTTCCCTG
1501 CCCCAGCCCC AGCAGCCATG GGTACATGGG TCCCCAGCTC ACCTATGGAT
1551 TCCCGCCAGT CTGCCAGCT GCAGTACTCA CGCCCATGGG GGGATCTTGG
1601 TCTGTTTTTC TTGTGGGAGC CTAGTGGAGA GCAGACGTGG CTTTTATGTA
1651 GTCTTGTTGG GGAGGTGACT TGCATGGTGG GGACAAGGCT GTCGTGGCAA
1701 CCTTGGGATC GAGTTTGAGA CTAAAGGATG TCATGAGATC CCTGGCTTCT
1751 CCCCATGTTG TTCCCGGACA AGGGCAGAAG GGAGGCATGG CAAGGGACCT
1801 CTGCTGTCTT TACTCAACAG TGGTCTCAT CCTTCCCCAC CTCCCCTGTC
1851 TTCTGCAAG GGCACCAAGT GTATGAGAAA GTTGGCCTTT GGACTTAGGA
1901 TTTCTTATTG TAGCTAAGAG CCATCTGAAG CAGCAGGTTG CAGGACAAAT
1951 GCTTCAGTCC GCCGAGAGCA GTACCGTGTG GCCAAGAGGT GGACTCAGAG
2001 CCTTCTTGA GCTAAACTCG GCCAACCAG GCACGCAGCA TGTCCCTCA
2051 GGTCTCCAGT CAGTCCAGGT TGACCTCAG TTCTGGACGT GTGTATATAG
2101 CTGATATTAA TACCTCAAGG TCATTGTGGC TCTGGGGATG CCAGGGCAGG
2151 AGGACGAGGG TGCGCTGTGG ACACAGCAGT CCGCGGAATT CCGTCTGGG
2201 AAGCCAATGG TCGCGGGCAC CCCTTGCTTC CTCCCTCTGT TGTCTGCTGT
2251 TGTGACACAC ATCAATGGCA ATAACCTCTT CCAACTCCTC GCAGAAGTGG
2301 GAGAGGCCGG CAGCCTGCAC CGAGAGGGGC TTTCTCTCTT CTGCTCCCC
2351 GCTTCGTTCT GTTTTGGCTG CAGAGAGTGG TTCATCCATA CTCTCATTCC
2401 CTCGCCCTCC CTTGTGGACG GGGGTCTTGC CTTTCAATT CCGTGTGTTT
2451 GGTGTCTTCC CTTATCTGCT ACCCTGAATC ACCTGTCCCT GTCTGTGCTG
2501 GTATGGGAA CATGCTTGTA AACTGCGTAA CAAATCTACT TTGTGTATGT
2551 GTCTGTTTAT GGGGGTGGTT TATTATTTTT GCTGGTCCCT AGACCACTTT
2601 GTATGACCGT TTGAGCTCTG AGCAGGCCAG GGGCTGACAG CTAATGTACG
2651 GACCTCAGC GGTGGAGCCT GCTGGGGGGA CCCAGCTGCT CTTGGACAAG
```

```

2701 TGGCTGAGCT CCTATCTGGC CTCCTCTTTT TTTTTTTTTT CAAGTAATTT
2751 GTGTGTATTT CTAACGTATT GTATTGAAAA AATTCCTAGT ATTCAGTAA
2801 AAATGCCTGT TGTGAGATGA ACCTCCTGTA ACTTCTATCT GTTCTTTTTT
2851 GAGGCTCAGG GAGAACTAG CATTTTTTTT TTCCAAACT ACTTTTGTG
2901 ACTGTGACAG TTGTAAATAA AGTTTGAAAA TGCTCAAAAA AAAAAAAAAA
2951 AAAAC

```

BLAST Results

Entry HSG19750 from database EMBL:
human STS A001X24.
Score = 1050, P = 1.9e-39, identities = 212/213

Entry HSG20267 from database EMBL:
human STS A005C12.
Score = 610, P = 4.1e-19, identities = 122/122

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 213 bp to 527 bp; peptide length: 105
Category: strong similarity to known protein
Classification: unset

```

1 MSIFYFPIHCP DYLSAKMTE VMMNTQPMEE IGLSPRKDGL SYQIFPDPSD
51 FDRRCKLKDR LPSIVVEPTE GEVESGELRW PEEFLVQED EQDNCEETAK
101 ENKEQ

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_lj9, frame 3

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8e-42

PIR:S52241 XLCL2 protein - African clawed frog, N = 1, Score = 443, P = 8.2e-42

>PIR:S52241 XLCL2 protein - African clawed frog
Length = 102

HSPs:

Score = 443 (66.5 bits), Expect = 8.0e-42, P = 8.0e-42
Identities = 80/104 (76%), Positives = 95/104 (91%)

```

Query: 1 MSIFYFPIHCPDYLSAKMTEVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR 60
MS+++PIHC DYLSA+MTEV+MNTQ M+EIGLSPRKD SYQIFPDPSDF+R CKLKDR
Sbjct: 1 MSVYFPIHCTDYLSAEMTEVIMNTQSMDEIGLSPRKD--SYQIFPDPSDFERCCKLKDR 58

Query: 61 LPSIVVEPTEGEVESGELRWPEEFLVQEDEQDNCEETAKENKE 104
LPSIVVEPTEG+VESGELRWPEEF+V ED++ C++T KEN++
Sbjct: 59 LPSIVVEPTEGDVESGELRWPEEFVVDKEDKGTCDQTKKENEQ 102

```

Pedant information for DKFZphfd2_lj9, frame 3

Report for DKFZphfd2_lj9.3

```

[LENGTH] 105
[MW] 12269.78
[pI] 4.40
[HOMOL] PIR:S52241 XLCL2 protein - African clawed frog 5e-44

```

[KW] Alpha_Beta

```
SEQ      MSYFPIHCPDYLRSAKMTVMMNTQPMEEIGLSPRKDGLSYQIFPDPSDFDRRCKLKDR
PRD      cccccccccchhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhc

SEQ      LPSIVVEPTGEVESEGELRWPPEEFLVQEDEQDNCEETAKENKEQ
PRD      cceeeccccccccccccccccccccccccccccchhhhhhhhhccc
```

(No Prosite data available for DKFZphfkd2_1j9.3)

(No Pfam data available for DKFZphfkd2_1j9.3)

DKFZphfkd2_24a15

group: transmembrane protein

DKFZphfkd2_24a15 encodes a novel amino acid protein with similarity to C. elegans cosmid R07G3.

The novel protein contains 1 transmembrane region.
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to C. elegans R07G3.8
membrane regions: 1
Summary DKFZphfkd2_24a15 encodes a novel 323 amino acid protein, with
similarity to C. elegans R07G3.8.

similarity to C. elegans R07G3.8

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1513 bp
Poly A stretch at pos. 1494, no polyadenylation signal found

```
1 GGGGTACTCG GCGGCGGCGG AGCGGGCGGC AGAGCAGGGC GCGGCGGACT
51 CGCAGGGTAC CACCATCTTA AGGACAGAAA AGCTACAGGA CTCTAGGAGG
101 CCACCGTCCT GATTTGGGAA GTCCAACCTA CTTTGGCCAG ACAGCAGCTA
151 AGCTGGTTCA TCCCATCAGC CTGGATTGGT GAAACTGAAT CACAGGAGAT
201 ATTTCCAGGT TTGCTGGGAT GGGAAACCTG CTCAAAGTCC TTACCAGGGA
251 AATTGAAAAC TATCCACACT TTTTCTGGA TTTTGAAAT GCTCAGCCTA
301 CAGAAGGAGA GAGAGAAATC TGGAACCAGA TCAGCGCCGT CCTTCAGGAT
351 TCTGAGAGCA TCCTTGACAG CCTGCAGGCT TACAAAGGCG CAGGCCCCAGA
401 GATCCGAGAT GCAATTCAAA ATCCCAATGA CATTGAGCTT CAAGAAAAAG
451 CTTGGAATGC GGTGTGCCCT CTTGTTGTGA GGCTAAAGAG ATTTTACGAG
501 TTTTCCATTA GACTAGAAAA AGCTCTTCAG AGTTTATTGG AATCTCTGAC
551 TTGTCCACCC TACACACCAA CCCAACACCT GGAAAGGGAA CAGGCCCTGG
601 CAAAGGAGTT TGCCGAAATT TTACATTTTA CCCTTCGATT CGATGAGCTG
651 AAGATGAGGA ACCCGGCTAT TCAGAATGAC TTCAGCTACT ACAGAAGAAC
701 AATCAGTCGC AACCGCATCA ACAACATGCA CCTAGACATT GAGAATGAAG
751 TCAATAATGA GATGGCCAAT CGAATGTCCC TCTTCTATGC AGAAGCCACG
801 CCAATGCTGA AAACCCTTAG CAATGCCACA ATGCACCTTG TCTCTGAAAA
851 CAAACTCTCT CCAATAGAGA ACACCACAGA CTGCCTCAGC ACAATGACAA
901 GTGTCTGTAA AGTCATGCTG GAAACTCCGG AGTACAGAAG TAGGTTTACG
951 AGTGAAGAGA CCCTGATGTT CTGCATGAGG GTGATGGTGG GAGTCATCAT
1001 CCTCTATGAC CATGTCCACC CTGTGGGAGC TTTCTGCAAG ACATCCAAGA
1051 TCGATATGAA AGGCTGCATA AAAGTTTGA AGGAGCAGGC CCCAGACAGT
1101 GTGGAGGGGC TGCTAAATGC CCTCAGGTTC ACTACAAAGC ACTTGAACGA
1151 TGAATCAACT TCCAAACAGA TTCGAGCAAT GCTTCAGTAG AGCTCTGCTC
1201 AAAGAAGAGG ATCTATGTGC TGACCTCAGA AGATGTATAT GTTTACATAA
1251 TTTAATACAG ATTGATGTTA ATACTTGTGT ATTTACATAA CCGTTTCCTT
1301 CTTGTCACTG AAATATATGG ACCTTAATTT GTATCCTGAC TGACTCAACC
1351 CAGCAGAGCA TAAATTGACT TGAGAGCCTT ACCTTTGATG TCTGAAATGA
1401 AACCCCTTC TCCAAAGGCA AAATTCGGAG ACTTTGATCT TTGCTACTGG
1451 AGTCCTTTAA CAACATCTAT AACGATAAAA AATTCCTAAT TGTCAAAAAA
1501 AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

WO 01/12659

PCT/IB00/01496

PS00005	314->317	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00007	231->240	TYR_PHOSPHO_SITE	PDOC00007
PS00008	297->303	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phfkd2_24a15.3)

DKFZphfkd2_24b15

group: metabolism

DKFZphfkd2_24b15 encodes a novel 612 amino acid protein with similarity to bacterial and yeast phosphoglucomutase and phosphomannomutases.

The novel protein contains a phosphoserine signature typical for phosphoglucomutase (EC 5.4.2.2) or phosphomannomutase (EC 5.4.2.8). Thus, the protein seems to be taking part in the conversion of hexose phosphates.

The new protein can find application in modulation of hexose metabolism pathways and as a new enzyme for biotechnologic production processes.

similarity to phosphomannomutases

complete cDNA, complete cds, EST hits
potential start at bp 30 matches kozak consensus PyCnatG,

Sequenced by GBF

Locus: map="158.8 cR from top of Chr4 linkage group"

Insert length: 2204 bp

Poly A stretch at pos. 2186, no polyadenylation signal found

```
1 GGGCTCTGCA GCGGTAGCAC AAGCTCAGCG ATGGCGGCTC CAGAAGGCAG
51 CGGTCTAGGC GAGGACGCCG GGCTGGACCA GGAGACCGCC CAGTGGCTGC
101 GCTGGGACAA GAATTCCTTA ACTTTGGAGG CAGTGAAACG ACTAATAGCA
151 GAAGGTAATA AAGAAGAACT ACGAAAATGT TTTGGGGCCC GAATGGAGTT
201 TGGGACAGCT GGCCTCCGAG CTGCTATGGG ACCTGGAATT TCTCGTATGA
251 ATGACTTGAC CATCATCCAG ACTACACAGG GATTTTGACG ATACCTGGAA
301 AAACAATTCA GTGACTTAAA GCAGAAAGGC ATCGTGATCA GTTTTGACGC
351 CCGAGCTCAT CCATCCAGTG GGGGTAGCAG CAGAAGGTTT GCCCGACTTG
401 CTGCAACCAC ATTTATCAGT CAGGGGATTC CTGTGTACCT CTTTCTGAT
451 ATAACGCCAA CCCCCTTTGT GCCCTTCACA GTATCACATT TGAACTTTG
501 TGCTGGAATC ATGATAACTG CATCTCACA TCCAAAGCAG GATAATGGTT
551 ATAAGGTCTA TTGGGATAAT GGAGCTCAGA TCATTTCCTC TCACGATAAA
601 GGGATTTCCTC AAGCTATTGA AGAAAACTCA GAACCGTGGC CTCAAGCTTG
651 GGACGATTCT TTAATTGATA GCAGTCCACT TCTCCACAAT CCGAGTGCTT
701 CCATCAATAA TGACTACTTT GAAGACCTTA AAAAGTACTG TTTCCACAGG
751 AGCGTGAAAC GGGAGACAAA GGTGAAGTTT GTGCACACCT CTGTCCATGG
801 GGTGGGTGAT AGCTTTGTGC AGTCAGCTTT CAAGGCTTTT GACCTTGTTT
851 CTCCTGAGGC TGTTCCTGAA CAGAGAGATC CGGATCCTGA GTTTCCAACA
901 GTGAAATACC CGAATCCCGA AGAGGGGAAA GGTGTCTTGA CTTTGTCTTT
951 TGCTTTGGCT GACAAAACCA AGGCCAGAAT TGTTTAGCT AACGACCCGG
1001 ATGCTGATAG ACTTGCTGTG GCAGAAAAGC AAGACAGTGG TGAATGGAGG
1051 GTGTTTTCAG GCAATGAGTT GGGGGCCCTC CTGGGCTGGT GGCTTTTAC
1101 ATCTTTGAAA GAGAAGAACC AGGATCGCAG TGCTCTCAAA GACACGTACA
1151 TGTGTGCCAG CACCGTCTCC TCCAAAATCT TGCGGGCCAT TGCTTTAAAG
1201 GAAGGTTTTC ATTTTGAGGA AACATTAAC TGGCTTAAGT GGATGGGAAA
1251 CAGAGCCAAA CAGCTAATAG ACCAGGGGAA AACTGTTTAA TTTGCATTTG
1301 AAGAAGCTAT TGGATACATG TGCTGCCCTT TTGTTCTGGA CAAAGATGGA
1351 GTCAGTGCCG CTGTCATAAG TGCAGAGTTG GCTAGCTTCC TAGCAACCAA
1401 GAATTTGTCT TTGTCTCAGC AACTAAAGGC CATTTATGTG GAGTATGGCT
1451 ACCATATTAC TAAAGCTTCC TATTTATCT GCCATGATCA AGAACCATT
1501 AAGAAATTAT TTGAAAACCT CAGAACTAC GATGGAAAAA ATAATTATCC
1551 AAAAGCTTGT GGCAAAATTG AAATTTCTGC CATTAGGGAC CTTACAACCTG
1601 GCTATGATGA TAGCCAACCT GATAAAAAAG CTGTTCTTCC CACTAGTAAA
1651 AGCAGCCAAA TGATCACCTT CACCTTTGCT AATGGAGGCG TGGCCACCAT
1701 GCGCACCAGT GGGACAGAGC CAAAATCAA GTACTATGCA GAGCTGTGTG
1751 CCCACCTGG GAACAGTGAT CCTGAGCAGC TGAAGAAGGA ACTGAATGAA
1801 CTGGTCAGTG CTATTGAAGA ACATTTTTC CAGCCACAGA AGTACAATCT
1851 GCAGCCAAAA GCAGACTAAA ATAGTCCAGC CTTGGGTATA CTTGCATTTA
1901 CTACAAATTA AGCTGGGTTT AACTTGTTAA GCAATATTTT TAAGGGCCAA
1951 ATGATTCAAA ACATCACAGG TATTTATGTG TTTTACAAAG ACCTACATTC
2001 CTCATTGTTT CATGTTTGAC CTTTAAGGTG AAAAAAGAAA ATGGCCAAAC
2051 CCAACAAACT AACATTCCTA CTAATAAGTT GAGCTTGGAC ATATTTTGAA
2101 TTTTGTGAAG TGAAGATTTT TAAACTGACT AACTTAAAAA AATAGATTGT
2151 AATTGATGTG CCTTAATTTG CATAAATCAT AAATGTAAAA AAAAAAATAA
2201 AAAA
```

BLAST Results

Entry HS705145 from database EMBL:

human STS WI-6820.
Score = 1261, P = 3.6e-52, identities = 253/254

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 31 bp to 1866 bp; peptide length: 612
Category: strong similarity to known protein

```

1 MAAPEGSGLG EDARLDQETA QWLRWDKNSL TLEAVKRLIA EGNKEELRKC
51 FGARMEFGTA GLRAAMGPGI SRMNDLTIIQ TTQGFRCYLE KQFSDLKQKG
101 IVISFDARAH PSSGGSSRRF ARLAATTFIS QGIPVYLFSD ITPTFPVPFT
151 VSHLKLKAGI MITASHNPKQ DNGYKVYWDN GAQIISPHDK GISQAIEENL
201 EPWPQAWDDS LIDSSPLLHN PSASINNDYF EDLKKYCFHR SVNRETKVKF
251 VHTSVHGVGH SFVQSAFKAF DLVPPEAVPE QRPDPPEFPT VKYPNPEEGK
301 GVLTLSFALA DKTARIVLA NDPDADRLAV AEKQDSGEWR VFSGNELGAL
351 LGWMLFTSWK EKNQDRSALK DTYMLSSTVS SKILRAIALK EGFHFEETLT
401 GFKWMGNRAK QLIDQKTVL FAFEEAIGYM CCPFVLDDKG VSAAVISAEI
451 ASFLATKNLS LSQQLKAIYV EYGYHITKAS YFICHQDQETI KKLLENLRNY
501 DGKNNYPKAC GKFEISAIRD LTTGYDDSQP DKKAVLPTSK SSQMITTFA
551 NGGVATMRTS GTEPKIKYYA ELCAPPGNSD PEQLKKELNE LVSAIEEHFF
601 QPQKYNLQPK AD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfk2_24b15, frame 1

TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B, N = 1, Score = 1431, P = 1.6e-146

TREMBL:SPCC1840_5 gene: "SPCC1840.05c"; product: "similarity to phosphomannomutases"; S.pombe chromosome III cosmid c1840., N = 1, Score = 1210, P = 4.2e-123

PIR:S54585 hypothetical protein YMR278w - yeast (Saccharomyces cerevisiae), N = 1, Score = 1046, P = 1e-105

PIR:A71299 probable phosphomannomutase (manB) - syphilis spirochete, N = 1, Score = 697, P = 9.7e-69

>TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B Length = 595

HSPs:

Score = 1431 (214.7 bits), Expect = 1.6e-146, P = 1.6e-146
Identities = 285/598 (47%), Positives = 393/598 (65%)

```

Query:   13 ARLDQETAQWLRWDKNSLTLEAVKRLIAEGNKEELRKC FGARMEFGTAGLRAAMGPGISR 72
          A+LD++ A WL WDKN      +++L+ E N + L+      R+ FG TAG+R+ M G R
Sbjct:   6 AKLDKQVADWLAWDKNDKNRNEIQKLVDEKNVDALKARMDTRLVFGTAGVRSFPMQAGFGR 65

Query:   73 MNDLTIIQTQGFRCYLEKQFSDLKQKGIVISFDARAH PSSGGSSRRFARLAATTFISQG 132
          +NDLTIIQ T GF R++   + K G+ I FD R +      SRRFA L+A F+
Sbjct:   66 LNDLTIIQITHGFARHMLNVYGQPKN-GVAIGFDGRYN-----SRRFAELSANVEVRNN 118

Query:   133 IPVYLFSDITPTFPVPFTVSHLKLKAGIMITASHNPKQDNGYKVYWDNGAQIISPHDKGI 192
          IPVYLF S+++PTP V +   L AG++ITASHNPK+DNGYK YW NGAQII PHD I
Sbjct:   119 IPVYLFSEVSPTPVVSWATIKLGCDAGLIITASHNPKEDNGYKAYWSNGAQIIGPHDTEI 178

Query:   193 SQAIEENLEFPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHRSVNRETKVKFVH 252
          + E +P + WD S + SSPL H+   I+ YFE K F R +N T +KF +
Sbjct:   179 VRKEAEPPQPRDEYWDLSSELKSSPLFHSADVVID-PYFEVEKSLNFTREINGSTPLKFTY 237

Query:   253 TSVHGVGHSEFVQSAFKAFDLVPPE--AVPEQRDPDPPEFTVKYPNPEEGKGVLTLSFALA 310
          ++ HG+G+ + + F F      +V EQ+DP+P+FPT+ +PNPEEG+ VLT L+ A
Sbjct:   238 SAFHGIGYHYTKRMFAEFGFPASSFISVAEQQDPNPDFPTIPFPNPEEGRKVLTLAMETA 297

```

Query: 311 DTKKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWLFTSWKEKNQDRSALK 370
 DK + ++LANDPDADR+ +AEKQ GEWRVF+GNE+GAL+ WW++T+W++ N + A K
 Sbjct: 298 DKNGSTVILANDPDADRIQMAEKQKDGGEWRVFTGNEMGALITWWIWTNWRKANPNADASK 357

Query: 371 DTYMLSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQGKTVLFAFEEAIGYM 430
 Y+L+S VSS+I++ IA EGF E TLTGFKWMGNRA++L G V+ A+EE+IGYM
 Sbjct: 358 -VYILNSAVSSQIVKTIADAEGFKNETTTLTGFKWMGNRAEELRADGNQVILAWESIGYM 416

Query: 431 CCP-FVLDDKDGVSAAVISAELASFLATKNLSLSQLKAIYVEYGYHITKASYFICHQDET 489
 P +DKDGVSA + AE+A+FL + SL QL A+Y YG+H+ +++Y++ E
 Sbjct: 417 --PGHTMDKDGVSAAVFAEIAAFLHAEGKSLQDQLYALYRNGFHLVRSTYWMVPAPEV 474

Query: 490 IKKLFENLRNYDGKNYPKACGKFEISAIRDLTTGYDDSQPKKAVLPTSKSSQMITFTF 549
 KKLF LR D K +P G+ E++++RDLT GYD+S+PD K VLP S SS+M+TF
 Sbjct: 475 TKKLFSTLRA-DLK--FPTKIGEAEVASVRDLTIGYDNSKPDNKPVLPLSTSEMVTFFL 531

Query: 550 ANGGVATMRSTGTEPKIKYYAELCAPPGNS--DPEQLKKELNELVSAIEEHFFQPKYNL 607
 G V T+R SGTEPKIKYY EL PG + D E + E+++L + +PQ++ L
 Sbjct: 532 KTGSVTTLRASGTEPKIKYYIELITAPGKTQNDLESVISEMDQLEKDVVATLLRPQQFGL 591

Query: 608 QPK 610
 P+
 Sbjct: 592 IPR 594

Pedant information for DKFZphfd2_24b15, frame 1

Report for DKFZphfd2_24b15.1

[LENGTH] 612
 [MW] 68311.58
 [pI] 6.28
 [HOMOL] TREMBL:CEY43F4B_5 gene: "Y43F4B.5"; Caenorhabditis elegans cosmid Y43F4B le-157

[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YMR278w] 1e-111
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI0740] 3e-66
 [FUNCAT] c energy conversion [M. genitalium, MG053] 4e-50
 [FUNCAT] m outer membrane and cell wall [H. influenzae, HI1463] 2e-04
 [BLOCKS] BL00607D cAMP phosphodiesterases class-II proteins
 [BLOCKS] BL00710 Phosphoglucomutase and phosphomannomutase phosphoserine signa
 [EC] 5.4.2.8 Phosphomannomutase 3e-56
 [EC] 5.4.2.2 Phosphoglucomutase 1e-09
 [PIRKW] isomerase 3e-56
 [PIRKW] intramolecular transferase 3e-56
 [SUPFAM] Methanobacterium thermoautotrophicum phosphomannomutase 1e-06
 [SUPFAM] probable phosphorylating protein ureC 9e-06
 [PROSITE] PGM_PMM 1
 [PROSITE] MYRISTYL 10
 [PROSITE] LIPOCALIN 2
 [PROSITE] CK2_PHOSPHO_SITE 9
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 8
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Phosphoglucomutase and phosphomannomutase phosphoserine
 [KW] Alpha_Beta

SEQ MAAPEGSGLGEDARLDQETAQWLRLWDKNSLTLEAVKRLIAEGNKEELRKCFGARMEFGTA
 PRD cccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhhhccchhhhhhhhhhhhhccccc

SEQ GLRAAMGPGISRMDLTIIQTQGFCRYLEKQFSDLKQKGVISFDARAHPSGGSSRRF
 PRD cccccccccccccceeeehhhhhhhhhhhhhcccccceeeecccccccccccccchhh

SEQ ARLAATTFISQGPVYLFSDITPTFPVPFTVSHLKLKAGIMITASHNPKQDNGYKYVDN
 PRD hhhhhhhhhccceeeecccccccccchhhhhhhcccccceeeecccccccccccccceeeec

SEQ GAQIISPDKGISQAIEENLEPWPQAWDDSLIDSSPLLHNPSASINNDYFEDLKKYCFHR
 PRD cccccccccchhhhhhhhhhhhhhhcccccceccccccccccccccccchhhhhhhhhhhhhcc

SEQ SVNRETQKVFVHTSVHGVGHSFVQSAFKAFLVPPPEAVPEQRDPDEFPVTKYPNPEEGK
 PRD cccccceeeecccccchhhhhhhhhhhcccccceccccccccccccccccccccccccch

SEQ GVLTLFALADTKARIVLANDPDADRLAVAQKQDSGEWRVFSGNELGALLGWLFTSWK
 PRD hhhhhhhhhhhccceeeecccccceeeecccccceeeecccccchhhhhhhhhhhhh

SEQ EKNQDRSALKDITYMLSSTVSSKILRAIALKEGFHFEETLTGFKWMGNRAKQLIDQGKTVL
 PRD hccccccccceeeehhhhhhhhhhhhhcccccceccccccccchhhhhhhhhhhhhccceee

```

SEQ  FAFEEAIGYMCCPFVLDKDGVSAAVISAE LASFLATKNLSLSQQLKAIYVEYGYHITKAS
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  YFICHQDETIIKKLFENLRNYDGKNYPKACGKFEISAI RDLTTGYDDSQPKKAVLPTSK
PRD  eeecccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ  SSQMITFTTFANGGVATMRTSGTEPKIKYYAELCAPPGNSDPEQLKKELNELVSAIEEHFF
PRD  ccceeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QPQKYNLQPKAD
PRD  cccccccccccc

```

Prosites for DKFZphfd2_24b15.1

PS00001	458->462	ASN_GLYCOSYLATION	PDOC00001
PS00002	7->11	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS00005	117->120	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	380->383	PKC_PHOSPHO_SITE	PDOC00005
PS00005	489->492	PKC_PHOSPHO_SITE	PDOC00005
PS00005	538->541	PKC_PHOSPHO_SITE	PDOC00005
PS00005	556->559	PKC_PHOSPHO_SITE	PDOC00005
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	343->347	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	528->532	CK2_PHOSPHO_SITE	PDOC00006
PS00006	560->564	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	61->67	MYRISTYL	PDOC00008
PS00008	100->106	MYRISTYL	PDOC00008
PS00008	159->165	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00008	257->263	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00008	348->354	MYRISTYL	PDOC00008
PS00008	440->446	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00710	159->174	PGM_PMM	PDOC00589
PS00213	346->358	LIPOCALIN	PDOC00187
PS00213	344->358	LIPOCALIN	PDOC00187

Pfam for DKFZphfd2_24b15.1

HMM_NAME	Phosphoglucosaminidase and phosphomannosidase phosphoserine		
HMM	*GvnVidIGQNGMMPTPMIYFaIRTYKhmcmggGIMITaSHNPGGPDnDN		
	G+ V +	++PTP + F +	H+++ +GIMITASHNP DN
Query	132	GIPVYLFs---DITPTPFVPFTVS---HLKLCAGIMITASHNP--KQ-DN	172
HMM	GIK*		
	G+K		
Query	173	GYK	175

DKFZphfkd2_24e23

group: kidney derived

DKFZphfkd2_24e23 encodes a novel 198 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, 1 EST hit,
many ATGs in front of the ORF

Sequenced by GBF

Locus: unknown

Insert length: 1723 bp

Poly A stretch at pos. 1695, no polyadenylation signal found

```
1 GGGGGATTTT CGATCATGAC AACGATAGCA ATTGATATAC CTTCAAAATA
51 CGTGTCCAGT GAGTGTGTGAT TGTGTGTGGT TTCTCTAGGA GACCGTGTTC
101 ATGCAACACA GCATTATTTC ACCGCCTTTA CCCAGCTTC TTCATACACA
151 TGCACCTGTC AAGGGCTCTT TGGCTGAAGA GAAGTTAGAA GTTCCAGAT
201 ATGGAGGGGT ATTTTCAGCA GATATGCCCA CCGCCATGGT TTTGTCAGCT
251 CTGTAGGGTG GTCTTGACCC CTGCTCACTG CTGGCATCAC CTGAGCCTAT
301 GGCAGATACC CAGTGTCTGCC CGCCACCATG TGAATTCATC AGCTCTGCAG
351 GCACAGACCT TGCAC TAGGA ATGGGCTGGG ACGCCACCCT CTGCCTCTTA
401 CCATTCACCT GGTTTGGCAA GTGTGTGGG ATCTGGAATC ACATGGATGA
451 GGAACCCGAT AATGGTGACG ACCGAGGTAG CAGGCGAACC ACTGGCCAGG
501 GCAGGAAGTG GGCAGCTCAC GGGACTATGG CTGCACCGCG GGTTCATACC
551 GACTACCATC CTGGAGGTGG GAGCGCATGC TCATCTGTAA AAGTCCGGTC
601 CCACGTTGGA CACACCGGGG TCTTCTTCTT TGTGACCAG GATCCTCTGG
651 CAGTGTCTTT AACAAAGCCAG AGTCTGATCC CACCGCTCAT AAAGCCAGGG
701 TTGTTGAAAG CTGGGGGCTT CCTCCTCCTC TGTGCGCAGC CCTCAGCAAA
751 CGGTACACAG CTGTGCTGTC TGCTGTACAC CGACTTGGTA TCATCCCATG
801 AACTGTCCCC CTTTCGTGCT CTGTGCTTAG GGCCTCTGA TGCCCCATCT
851 GCCTGCGCTT CCTGCAACTG TTTAGCAAGC ACCTATTATC TATAGGGTGC
901 TGGGGTGCTG GGCAGGGCCA ATCGCTCCTA TTACTTTCTG CCCTGGGGAC
951 GTCTCTGTTT CCCACCTACC CCTGTAACGC CTCTGCTCTG CCTTCCCATC
1001 TGCGGGGCTA ACGCCATCCC ACAAGGGCTG GGCTGTCCGT TCAGAAGAGA
1051 AACTGGGAAG GGGCCTTGAG GACCTGTGTC CAGGCAGGGT GGACAAGGGC
1101 TTTGTGCAGG GAGCTCCTCT CCCATCTTTG TGTCTGACA GCCGTGACCG
1151 TGACCCCTCA AAGCAGAGCC AGTAGTGATC AGTATCCTGC TGCTTCAAGC
1201 CTGCACGGTC CTCTTCTCCT CTCCGCACAT CTGCATGCCT GTCAAACCCA
1251 GAGTAGTTTG GGGCCTGGTA AACAGAGGGA AGTTGGCTGG AGGAGGCCAG
1301 TCAGGAGTGC AAGAACCCCG CGTACTCTGT CCCACGTGGA TAAAGTCTCT
1351 AATTCCAGTC TGAGGTGAAT TCTTAGAGAG TGCTTTCATT TAATGTTTGC
1401 TTTATGCATT TCCCCTGCAG CTGTGACTAA TTGTGGAACA GCATACATT
1451 TGTTTTGAGA CTCTCTTGAG ATTTTCTTGG CAGTGTAAGG TCTACACCAT
1501 TTTCTCTCA GCATCAGAGA AGGCAGAAAG CAAGAGAAAG GAATGCAATG
1551 TGAGCAAGGC CAGGCACACT TGTGCTACTG CAGTTGGCAA GAATGGAGTC
1601 TAATCCAGC ACTTTGGGAG GCCGAGGCGG GTGGATCACC TGAGGTCAGG
1651 AATTTGAGAC CAACCTGGCC AACATGTTGA AACCTCGTCT GTACTAAAAA
1701 TACAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 299 bp to 892 bp; peptide length: 198
Category: putative protein

```

1 MADTQCCPPP CEFISSAGTD LALGMGWDAT LCLLPFTGFG KCAGIWNHMD
51 EEPDNGDDRG SRRTGQGRK WAAHGTMAAP RVHTDYHPGG GSACSSVKVR
101 SHVGHTGVFF FVDQDPLAVS LTSQSLIPPL IKPGLLKAWG FLLCAQPSA
151 NGHSLCCLLY TDLVSSHELSPFRALCLGPS DAPSACASCN CLASTYYL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfk2_24e23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfk2_24e23, frame 2

Report for DKFZphfk2_24e23.2

```

[LENGTH]      198
[MW]           20948.98
[pI]           6.01
[PROSITE]      MYRISTYL      5
[PROSITE]      AMIDATION     1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 2
[KW]           All_Beta
[KW]           LOW_COMPLEXITY 6.06 %

SEQ  MADTQCCPPPCFEISSAGTDLALGMGWDATLCLLPFTGFGKCAGIWNHMDDEPDNGDDRG
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SRRTGQGRKWAHGTMAAPRVHTDYHPGGGSACSSVKVRSHVGHTGVFFFVDQDPLAVS
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LTSQSLIPPLIKPGLLKAWGFLLLCAQPSANGHSLCCLLYTDLVSSHELSPFRALCLGPS
SEG  .....xxxxxxxxxxxx.....
PRD  eccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccc

SEQ  DAPSACASCNCLASTYYL
SEG  .....
PRD  cccccccccccccccccccccc

```

Prosites for DKFZphfk2_24e23.2

PS00004	62->66	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	96->99	PKC_PHOSPHO_SITE	PDOC00005
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00008	18->24	MYRISTYL	PDOC00008
PS00008	60->66	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008
PS00008	91->97	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKFZphfk2_24e23.2)

DKFZphfkd2_24n20

group: intracellular transport and trafficking

DKFZphfkd2_24n20.3 encodes a novel 366 amino acid protein with similarity to human eps8 binding protein e3B1 and spectrins.

The new protein contains an Src homology domain 3 and is similar to human eps8 SH3 domain binding protein 1 (e3B1) and spectrins. Eps8 is a substrate of receptor tyrosine kinases involved in mitogenic signaling. Spectrin is part of the submembrane cytoskeletal network in the human erythrocyte ghost. Nonerythroid spectrins are proposed to have roles in cell adhesion, establishment of cell polarity, and attachment of other cytoskeletal structures to the plasma membrane. The new protein seems to be part of the signalling pathway between tyrosine kinases and the membrane/cyto skeleton.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamics.

strong similarity to eps8 binding protein e3B1

complete cDNA, complete cds, few EST hits
potential start at Bp 300, but there are ATGs in other frames in
5' region of the cDNA

Sequenced by GBF

Locus: /map="17"

Insert length: 1719 bp

Poly A stretch at pos. 1699, polyadenylation signal at pos. 1680

```
1 GGGGACAGCT GCCCGACCT TGGCTTCCTC TGCTGGGTGG GATTGGGGGC
51 TGGGCCCCCA AATGGGCCCC TGGCTTCCTC CTTCCTCTGG GCAGGGGACA
101 GAGAGACACA GGCTCGGGGA GCAGGACTGA CTTCCTCTTG TCCCGGAATG
151 AGCATGCGTG CCCTTTGCAA GCAGGTTTGG GTCTCACGCA GAGGAAACCA
201 AAAGCAATAA GAGGGAGGGA AGGCAGAGCA ACCAATCAAG GGCAGGGTGA
251 GACTCAAAAC GAGCGGGCTC CCTGGGGAGC CAGACAGAGG CTGGGGGTGA
301 TGGCGGAGCT ACAGCAGCTG CAGGAGTTTG AGATCCCCAC TGGCCGGGAG
351 GCTCTGAGGG GCAACCACAG TGCCCTGCTG CGGGTCGCTG ACTACTGCGA
401 GGACAACATAT GTGCAGGCA CAGACAAGCA GAAGGCGCTG GAGGAGACCA
451 TGGCCTTCAC TACCCAGGCA CTGGCCAGCG TGGCCTACCA GGTGGGCAAC
501 CTGGCCGGGC AACTCTGCG CATGTTGGAC CTGCAGGGGG CCGCCCTGCG
551 GCAGGTGGAA GCCCGTGTA GCACGCTGGG CCAGATGGTG AACATGCATA
601 TGGAGAAGGT GGCCCGAAGG GAGATCGGCA CCTTAGCCAC TGTCAGCGG
651 CTGCCCCCGG GCCAGAAGGT CATCGCCCCA GAGAACCTAC CCCCTCTCAC
701 GCGCCTACTGC AGGAGACCCC TCAACTTTGG CTGCCTGGAC GACATTGGCC
751 ATGGGATCAA GGACCTCAGC ACGCAGCTGT CAAGAACAGG CACCCTGTCT
801 CGAAGAGCA TCAAGGCCCC TGCCACACCC GCCTCCGCCA CCTTGGGGAG
851 ACCGCCCCGG ATTCCCGAGC CAGTGCACCT GCCGGTGGTG CCGCAGGGCA
901 GACTCTCGCG CGCCTCCTCT GCGCTTTCCT TGGCCTCGGC CGGCAGCGCC
951 GAAGGTGTCT GTGGGGCCCC CACGCCCAAG GGGCAGGCAG CACCTCCAGC
1001 CCCACCTCTC CCCAGCTCCT TGGACCCACC TCCTCCACCA GCAGCCGTCG
1051 AGGTGTTCCA GCGGCCTCCC ACGCTGGAGG AGTTGTCCCC ACCCCACCG
1101 GACGAAGAGC TGCCCTGCC ACTGGACCTG CCTCCTCTC CACCCCTGGA
1151 TGGAGATGAA TTGGGGCTGC CTCCACCCCC ACCAGGATT GGGCCTGATG
1201 AGCCAGCTG GGTGCCTGCC TCATACTTGG AGAAAGTGGT GACACTGTAC
1251 CCATACACCA GCCAGAAGGA CAATGAGCTC TCCTTCTCTG AGGGCACTGT
1301 CATCTGTGTC ACTCGCGCT ACTCCGATGG CTGGTGCGAG GCGCTCAGT
1351 CGGAGGGGAC TGGATTCTTC CCTGGGAAT ATGTGGAGCC CAGCTGCTGA
1401 CAGCCAGGG CTCTCTGGGC AGCTGATGTC TGCACTGAGT GGGTTTCATG
1451 AGCCCCAAGC CAAAACACGC TCCAGTCACA GCTGGACTGG GTCTGCCAC
1501 CTCTTGGGCT GTGAGCTGTG TTCTGTCTT CCTCCATCG GAGGGAGAAG
1551 GGGTCTTGGG GAGAGAGAAT TTATCCAGAG GCCTGCTGCA GATGGGGAAG
1601 AGCTGGAAC CAAGAAGTTT GTCAACAGAG GACCCTACT CCATGCAGGA
1651 CAGGTCTCC TGCTGCAAGT CCCAAGTTTG AATAAACAG ATGATGTCCA
1701 AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

Entry AC004797 from database EMBL:
Homo sapiens chromosome 17, clone hRPC.62 O.9, complete sequence.
Score = 2316, P = 5.9e-255, identities = 464/465
7 exons Bp 93317-110902

Medline entries

97163405:
Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth.

98256293:
Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton.

Peptide information for frame 3

ORF from 300 bp to 1397 bp; peptide length: 366
Category: strong similarity to known protein

```

1 MAELQQLQEF EIPTGREALR GNHSALLRVA DYCEDNYVQA TDKQKALEET
51 MAFTTQALAS VAYQVGNLAG HTLRMLDLQG AALRQVEARV STLGMVNMH
101 MEKVARREIG TLATVQRLPP GQKVIAPENL PPLTPYCRRP LNFGCLDDIG
151 HGIKDLSTQL SRTGTLSRKS IKAPATPASA TLGRPPRIPE PVHLPVVVPDG
201 RLSAASSASS LASAGSAEGV GGAPTPKGQA APPAPPLPSS LDPPPPPAAV
251 EVFQRPPTLE ELSPPPPDEE LPLPLDLPPP PPLDGDDELGL PPPPPGFGPD
301 EPSWVPASYL EKVVTLYPYT SQKDNELSFS EGTVICVTRR YSDGWCEGVS
351 SEGTFGFFGN YVEPSC

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfk2_24n20, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphfk2_24n20, frame 3

Report for DKFZphfk2_24n20.3

```

[LENGTH]      366
[MW]           38947.21
[pI]           4.93
[COMOL]        TREMBL:U87166.1 gene: "SSH3BP1"; product: "spectrin SH3 domain binding protein
1"; Homo sapiens spectrin SH3 domain binding protein 1 (SSH3BP1) mRNA, complete cds. 3e-48

[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YGR136w] 9e-06
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YPR154w] 3e-05
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YDR388w] 2e-04
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR388w]
2e-04
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YDR162c] 4e-04
[BLOCKS]       BL50002B Src homology 3 (SH3) domain proteins profile
[SUPFAM]       SH3 homology 6e-17
[PROSITE]      MYRISTYL 6
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 6
[PROSITE]      PKC_PHOSPHO_SITE 8
[PROSITE]      ASN_GLYCOSYLATION 1
[PFAM]         Src homology domain 3
[KW]           Irregular
[KW]           3D
[KW]           LOW_COMPLEXITY 24.04 %

```

```

SEQ  MAELQQLQEF EIPTGREALRGNHSALLRVADYCEDNYVQATDKQKALEETMAFTTQALAS
SEG  .....
laboA .....

SEQ  VAYQVGNLAGHTLRMLDLQGAALRQVEARVSTLGMVNMHMEKVARREIGTLATVQRLPP
SEG  .....
laboA .....

```

```

SEQ      GQKVIAPENLPPLTPYCRRLNFGCLDDIGHGIKDLSTQLSRTGTLRSKSIKAPATPASA
SEG      .....
laboA    .....

SEQ      TLGRPPRIPEPVHLPVVPDGRLSAASSASSLASAGSAEGVGGAPTPKGQAAPPAPPLPSS
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
laboA    .....

SEQ      LDPPPPPAAVEVFQRPPTLEELSPPPPDEELFLPLDLPPPPPLDGDDELGLPPPPPGFGPD
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
laboA    .....

SEQ      EPSWVPASYLEKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSEGTGFFPGN
SEG      xx.....
laboA    .....EECCCBCCCTTTBCCBTTEEEEEETTTTEEEEEETEEEEEGG

SEQ      YVEPSC
SEG      .....
laboA    GEEE..

```

Prosites for DKFZphfkd2_24n20.3

PS00001	22->26	ASN_GLYCOSYLATION	PDOC00001
PS00004	339->343	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	14->17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	72->75	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	225->228	PKC_PHOSPHO_SITE	PDOC00005
PS00005	321->324	PKC_PHOSPHO_SITE	PDOC00005
PS00005	338->341	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	321->325	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00008	21->27	MYRISTYL	PDOC00008
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	94->100	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	215->221	MYRISTYL	PDOC00008
PS00008	332->338	MYRISTYL	PDOC00008

Pfam for DKFZphfkd2_24n20.3

```

HMM_NAME      Src homology domain 3
HMM            *pyVIALYDYqAqdpDELSFkEGDIIiIIEdsDD.WWrgRnnnTNGQEGW
               ++V+ LY+Y++Q ++ELSF EG +I + + D W++G + +G+
Query          311 EKVVTLYPYTSQKDNELSFSEGTVICVTRRYSDGWCEGVSSE---GTGF 356

HMM            IPSNYVEPi*
               +P NYVEP
Query          357 FPGNYVEPS 365

```

DKFZphfkd2_24p5

group: intracellular transport and trafficking

DKFZphfkd2_24p5 encodes a novel 811 amino acid protein which is a novel splice variant of human ankyrin G.

The ankyrin 3 gene encodes a novel ankyrin, which is expressed in multiple tissues, with very high expression at the axonal initial segment and nodes of Ranvier of neurons in the central and peripheral nervous systems. Ankyrin G shows several tissue-specific alternative mRNA processing. The different ankyrin G proteins participate in maintenance/targeting of ion channels and cell adhesion molecules to nodes of Ranvier and axonal initial segments.

The new protein can find application in modulating the structure and membrane topology of Ranvier nodes and other neuronal cell membranes.

Human ankyrin G (ANK-3) new splice variant

splice variant
potential frame shift at 2720 was checked
see BLASTX

Sequenced by EMBL

Locus: /map="10q21"

Insert length: 3470 bp
Poly A stretch at pos. 3459, no polyadenylation signal found

```
1 AGCTTTAAAA GGATGTCTGC GAAGTGGTCA AAAGGATCTT AACCTCAATT
51 AAGTGGGGTT TTTTAAAAAG ATTTTGTGG GGCCTGAAA TTTTGAAAAT
101 CTTTGAACCTC TGAGTGGGGA AAGATGTATA ATTCTCAAT TGCCTACGAG
151 GATATCAAGA TGCTGAGAGG AATTCAGCGG TGGTGAAGAG AGTGGATACA
201 AACCAGGGAT TGGTTTCCTT GAGCTGTTT GGAGGTGAT TCTAAATCAC
251 TGCTTAAGGA ATTCCTGGAA ACATCAGGAA AACATTGAT CATCCAAGCC
301 TAGTGGAAAT GGCTTTACCG CAGAGTGAAG ATGCAATGAC CGGGGACACA
351 GACAAATATC TTGGGCCACA GGACCTTAAG GAATTGGGTG ATGATTCCCT
401 GCCTGCAGAG GGTACATGG GCTTTAGTCT CGGAGCGCGT TCTGCCAGCC
451 TCCGCTCCTT CAGTTCGGAT GGGTCTTACA CCTTGAACAG AAGCTCCTAT
501 GGACGGGACA GCATGATGAT TGAAGAACTC CTCGTGCCAT CCAAAGAGCA
551 GCATCTAACA TTCACAAGGG AATTGTATC AGATTCTCTT AGACATTACA
601 GCTGGGCTGC AGACACCTTA GACAATGTCA ATCTTGTTCC AAGCCCCATT
651 CATTCTGGGT TTCTGGTTAG CTTTATGGTG GACGCGAGAG GGGGCTCCAT
701 GAGAGGAAGC CGTCATCACG GGATGAGAAT CATCATTCCT CCACGCAAGT
751 GTACGGCCCC CACTCGAATC ACCTGCCGTT TGGTAAAGAG ACATAAACTG
801 GCCAACCCAC CCCCCATGGT GGAAGGAGAG GGATTAGCCA GTAGGCTGGT
851 AGAAATGGGT CCTGCAGGGG CACAATTTT AGGCCCTGTG ATAGTGGAAA
901 TCCCTCACTT TGGGTCCATG AGAGGAAAAG AGAGAGAAT CATTGTTCTT
951 CGAAGTGAAA ATGGTGAAAC TTGGAAGGAG CATCAGTTTG ACAGCAAAAA
1001 TGAAGATTTA ACCGAGTTAC TTAATGGCAT GGATGAAGAA CTTGATAGCC
1051 CAGAAGAGTT AGGGAAAAAG CGTATCTGCA GGATTATCAC GAAAGATTTC
1101 CCCCAGTATT TTGCAGTGGT TTCCCGGATT AAGCAGGAAA GCAACCAGAT
1151 TGGTCTCGAA GGTGGAATTC TGAGCAGCAC CACAGTGCCC CTTGTTCAAG
1201 CATCTTTCCC AGAGGGTGCC CTAACTAATA GAATTCGAGT GGCCTCCAG
1251 GCCCAGCCTG TTCCAGATGA AATTGTGAAA AAGATCCTTG GAAACAAAGC
1301 AACTTTTAGC CCAATTGTCA CTGTGGAACC AAGAAGACGG AAATTCCTAT
1351 AACCAATCAC AATGACCATT CCGGTGCCCC CGCCCTCAGG AGAAGGTGTA
1401 TCCAATGGAT ACAAAGGGGA CACTACACCC AATCTGCGTC TTCTCTGTAG
1451 CATTACAGGG GGCACCTTCG CTGCTCAGTG GGAAGACATC ACAGGAACAA
1501 CTCCTTTGAC GTTTATAAAA GATTGTGTCT CCTTTACAAC CAATGTTTCA
1551 GCCAGATTTT GGCTTGCAAG CTGCCATCAA GTTTTAGAAA CTGTGGGGTT
1601 AGCCACGCAA CTGTACAGAG AATTGATATG TGTTCATAT ATGGCCAAGT
1651 TTGTTGTTTT TGCCAAAATG AATGATCCCG TAGAATCTTC CTTGCGATGT
1701 TTCTGCATGA CAGATGACAA AGTGGACAAA ACTTTAGAGC AACAAGAGAA
1751 TTGTGAGGAA GTCGCAAGAA GCAAAGATAT TGAGGTCTCT GAAGGAAAC
1801 CTATTTATGT TGATTGTTAT GGAAATTGG CCCCCTTAC CAAGGAGGA
1851 CAGCAACTTG TTTTAACTT TTATCTTTC AAAGAAAATA GACTGCCATT
1901 TTCCATCAAG ATTAGAGACA CCAGCCAAGA GCCCTGTGGT CGTCTGTCTT
1951 TTCTGAAAGA ACCAAAGACA ACAAAGGAC TGCCCTCAAC AGCGGTTTGC
2001 AACTTAAATA TCACTCTGCC AGCACATAAA AAGATTGAGA AACAGATGG
2051 ACGACAGAGC TTCCATCCTT TAGCTTTACG TAAGCGCTAC AGCTACTTGA
2101 CTGAGCCTGG AATGAGTCCA CAGAGTCCAT GTGAACGGAG AGATATCAGG
2151 ATGGCAATAG TAGCCGATCA CCTGGGACTT AGTTGGACAG AACTGGCAAG
2201 GGAACCTGAT TTTTCACTGG ATGAAATCAA TCAAAATACG GTGGAAAATC
2251 CAAATCTCTT AATTCTCAG AGCTTCATGT TTTTAAAAAA ATGGGTTACC
2301 AGAGACGGAA AAAATGCCAC AACTGATGCC TTAACCTCGG TCTTGACAAA
2351 AATTAATCGA ATAGATATAG TGACACTGCT AGAAGGACCA ATATTGATT
```

```

2401 ATGGAATAT TTCAGGCACC AGAAGTTTG CAGATGAGAA CAATGTTTTC
2451 CATGACCCCTG TTGATGGTTA TCCTTCCCTT CAAGTGGAAAC TGGAAACCCC
2501 CACAGGGTTG CACTACACAC CACCTACCCC TTCCAGCAA GATGATTATT
2551 TTAGTGATAT CTCTAGCATA GAATCTCCCC TAGAACCCC TAGTAGACTG
2601 AGTGATGGGC TAGTGCCCTC CCAGGGGAAC ATAGAGCATT CCGCAGATGG
2651 ACCTCCAGTC GTAACGTCAG AAGACGCTTC CTTAGAAGAC AGCAAACTGG
2701 AAGACTCAGT GCCTTTAACA GAAATGCCTG AAGCAGTGAT GTAGATGAGA
2751 GCCAGTTGGA GAATGTATGT CTGAGTTGGC AGAATGAGAC ATCAAGTGGA
2801 AACCTAGAGT CCTGCGCTCA AGCTCGAAGA GTAACCTGGT GGTACTAGA
2851 TCGACTGGAT GACAGCCCTG ACCAGTGTAG AGATTCCATT ACCTCATATC
2901 TCAAAGGAGA AGCTGGCAAA TTTGAAGCAA ATGGAAGCCA TACAGAAATC
2951 ACTCCAGAAG CAAAGACAAA ATCTTACTTT CCAGAATCCC AAAATGATGT
3001 AGGAAAACAG AGTACCAAGG AAACCTCGAA ACCAAAAATA CATGGATCTG
3051 GTCATGTTGA AGAACCAGCA TCACCACTAG CAGCATATCA GAAATCTCTA
3101 GAAGAAACCA GCAAGCTTAT AATAGAAGAG ACTAAACCTT GTGTGCCTGT
3151 CAGTATGAAA AAGATGAGTA GGACTTCTCC AGCAGATGGC AAGCCAAGGC
3201 TTAGCTCCA TGAAGAAGAG GGTCCAGTG GGTCTGAGCA AAGCAGGGA
3251 GAAGGTTTTA AGGTGAAAC GAAGAAAGAA ATCCGGCATG TGGAAAAGAA
3301 GAGCCACTCG TAACAGCGAA CGGTCAGTCA AGGATCATAA GTTTTACTG
3351 CCAGTATTGA GAAATTCGTG GAAGAAATGT CAGCAGGAAG TAAAAATCA
3401 CCGAGAAGTG TGTGTGTGTT CGCTGCTTCC ACACATAAT GGCATGATT
3451 TTTTATGCA AAAAAAAAAA

```

BLAST Results

```

-----
Entry MMANK3A_1 from database TREMBL:
Ank3"; product: "ankyrin 3"; Mus mu... +3 4022 0.0 2

Entry HS13616 from database EMBL:
Human ankyrin G (ANK-3) mRNA, complete cds.
Length = 14,770
Plus Strand HSPs:
Score = 8505 (1276.1 bits), Expect = 0.0, Sum P(3) = 0.0
Identities = 1799/1873 (96%)

```

Medline entries

```

-----
95394457:
Chromosomal localization of the ankyrinG gene
(ANK3/Ank3) to human 10q21 and mouse 10.

95138209:
A new ankyrin gene with neural-specific isoforms localized at the
axonal initial segment and node of Ranvier

```

Peptide information for frame 3

```

-----
ORF from 309 bp to 2741 bp; peptide length: 811
Category: known protein
Classification: unset

```

```

1 MALPQSEDAM TGDYDKYLG P QDLKELGDDS LPAEGYMGFS LGARSASLRS
51 FSSDGSYTLN RSSYARDMM IEELLVPSKE QHLTFTRFED SDLSLRHYSWA
101 ADTLNVDNLV PSPIHSGFLV SFMVDARGGS MRGSRHHGMR IIPPRKCTA
151 PTRITCRLVK RHKLANPPP VEGEGLASRL VEMGPAGAOF LGPVIVEIPH
201 FGSMRGKERE LIVLRSENGE TWKEHQFDSK NEDLTELLNG MDEELDSPEE
251 LGKKRICRII TKDFPQYFAV VSRIKQESNQ IGPEGGILSS TTVPLVQASF
301 PEGALTKRIR VGLQAQPVDP EIVKKILGNK ATFSPIVTVE PRRRKFKKPI
351 TMTIPVPPPS GEGVSNYKYG DTTPNLRLLC SITGGTSPAQ WEDITGTTPL
401 TFIKDCVSFT TNVSARFWLA DCHQVLETvG LATQLYRELI CVPYMAKFV
451 FAKMNDPVES SLRCFCMTDD KVDKTLQQE NFEEVARSKD IEVLEGKPIY
501 VDCYGNLAPL TKGGQQLVFN FYSFKNRPL FSIKIRDTSQ EPCGRLSFLK
551 EPKTTKGLPQ TAVCNLNLITL PAHKKIEKTD GRQSFASLAL RKRYSYLTEP
601 GMSQSPCER TDIRMAIVAD HLGLSWTELA RELNFSVDEI NQIRVENPNS
651 LISQSFMLK KWTVDGKNA TTDALTSVLT KINRIDIVTL LEGPIFDYGN
701 ISGTRSFADN NNVFHDPVDG YPSLQVELET PTGLHYTPPT PFQDDYFSD
751 ISSIESPLRT PSRLSDGLVP SQGNIEHSAD GPPVVTAEDE SLEDSKLEDS
801 VPLTEMPEAV M

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfkd2_24p5, frame 3

TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds., N = 1, Score = 4022, P = 0

TREMBL:MMANK3B_3 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

TREMBL:MMANK3B_4 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (7kb isoform) mRNA, complete cds., N = 1, Score = 4005, P = 0

>TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds.
Length = 1,094

HSPs:

Score = 4022 (603.5 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 769/805 (95%), Positives = 783/805 (97%)

```
Query:      1 MALPQSEDAMTGDTKYLGPDQLKELGDDSLPAEGYMGFSLGARSASLSRFSDDGSYTLN 60
             MALP SEDA+TGOTDKYLGPDQLKELGDDSLPAEGY+GFSLGARSASLSRFSDD SYTLN
Sbjct:      1 MALPHSEDAITGDTDKYLGPDQLKELGDDSLPAEGYVGFSLGARSASLSRFSDDSYTLN 60

Query:      61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVPSPIHSGFLV 120
             RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLV SP+HSGFLV
Sbjct:      61 RSSYARDSMMIEELLVPSKEQHLTFTREFDSDSLRHYSWAADTLDNVNLVSSPVHSGFLV 120

Query:      121 SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180
             SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL
Sbjct:      121 SFMVDARGGSMRGSRRHGMRIIIPPRKCTAPTRITCRLVKRHKLANPPPMVEGEGLASRL 180

Query:      181 VEMGPAGAQLGPGVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLLELLNG 240
             VEMGPAGAQLGPGVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDL ELLNG
Sbjct:      181 VEMGPAGAQLGPGVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLAELLNG 240

Query:      241 MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300
             MDEELDSPEELG KRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF
Sbjct:      241 MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGGILSSTTVPLVQASF 300

Query:      301 PEGALTKRIRVGLQAQVPVDEIVKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS 360
             PEGALTKRIRVGLQAQVPV+E VKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS
Sbjct:      301 PEGALTKRIRVGLQAQVPVDEIVKKILGNKATFSPIVTEPRRRKFHKPITMTIPVPPPS 360

Query:      361 GEGVSNYKGDTPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420
             GEGVSNYKGD TPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
Sbjct:      361 GEGVSNYKGDATPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA 420

Query:      421 DCHQVLETVGLATQLYRELICVPYMAKFVVFVAKMNDPVESLRCFCMTDDKVDKLEQQE 480
             DCHQVLETVGLA+QLYRELICVPYMAKFVVFVAK NDPVESLRCFCMTDD+VDKLEQQE
Sbjct:      421 DCHQVLETVGLASQLYRELICVPYMAKFVVFVAKTNDPVESLRCFCMTDDRVDKLEQQE 480

Query:      481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ 540
             NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ
Sbjct:      481 NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYFSENRLPFSIKIRDTSQ 540

Query:      541 EPCGRLSFLKEPKTTKGLPQTAVCNLNLITLPAHKKIEKTDGRQSFASLALRKRYSYLTP 600
             EPCGRLSFLKEPKTTKGLPQTAVCNLNLITLPAHKK EK D RQSFASLALRKRYSYLTP
Sbjct:      541 EPCGRLSFLKEPKTTKGLPQTAVCNLNLITLPAHKKAEKADRRQSFASLALRKRYSYLTP 600

Query:      601 GMSQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLK 660
             GMSQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSF LK
Sbjct:      601 GMSQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNLSISQSFMLLK 660

Query:      661 KVVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG 720
             KVVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG
Sbjct:      661 KVVTRDGKNATTDALTSVLTKINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG 720

Query:      721 YPSLQVELETPTGLHYTPPTPFQDDYFSDISSIESPLRTPSRSLSDGLVPSQGNIEHSAD 780
             +PS QVELET P GL++TTP PFQDD+FSDISSIESP RTPSRSLSDGLVPSQGNIEH
Sbjct:      721 HPSFQVELETFMGLYWTTPNPFQDDHFSDISSIESPFRTPSRSLSDGLVPSQGNIEHPTG 780

Query:      781 GPPVVTAEDASLEDSKLEDSVPLTE 805
             GPPVVTAED SLEDSK++DSV +T+
```

Sbjct: 781 GPPVVTAEDTSLSDSKMDDSVTVTD 805

Pedant information for DKFZphfkd2_24p5, frame 3

Report for DKFZphfkd2_24p5.3

[LENGTH] 811
 [MW] 90104.66
 [pI] 5.40
 [HOMOL] TREMBL:MMANK3A_1 gene: "Ank3"; product: "ankyrin 3"; Mus musculus epithelial
 ankyrin 3 (Ank3) 5kb isoform mRNA, complete cds. 0.0
 [BLOCKS] BL50017B Death domain proteins profile
 [PIRKW] phosphoprotein 0.0
 [PIRKW] alternative splicing 0.0
 [PIRKW] peripheral membrane protein 0.0
 [PIRKW] cytoskeleton 0.0
 [SUPFAM] ankyrin 0.0
 [SUPFAM] ankyrin repeat homology 0.0
 [SUPFAM] unassigned ankyrin repeat proteins 0.0
 [KW] TRANSMEMBRANE 2
 [KW] LOW_COMPLEXITY 1.73 %

SEQ MALPQSEDAMTGDTKYLGPQDLKELGDDSLPAEGYMGFSLGARSASLSRFSSDGSYTLN
 SEG
 PRD ccc
 MEM
 SEQ RSSYARDSMMIEELLVPSKEQHLTFTRFSDSLRHYSWAADTLDNVNLVPSPIHSGFLV
 SEG
 PRD cccchhhhhhhhhheeehhhhhhhhhhccccccccccccccccccccccccccccccccccc
 MEMMMMMMMMMMMMM
 SEQ SFMVDARGGSMRSGRHGMRIIPPRKCTAPTRITCRLVKRHLKLANPPPMVEGEGLASRL
 SEGxxxxxxxxxxxxxxxx
 PRD eeeeeccc
 MEM MM
 SEQ VEMGPAGAQLGFPVIVEIPHFGSMRGKERELIVLRSENGETWKEHQFDSKNEDLTELLNG
 SEG
 PRD eeeeeccc
 MEM MM
 SEQ MDEELDSPEELGKKRICRIITKDFPQYFAVVSRIKQESNQIGPEGILSSTTVPLVQASF
 SEG
 PRD cccccchhhhhhhhhheeecc
 MEM
 SEQ PEGALTKRIRVGLQAQVPDEIVKKILGNKATFSPIVTVEPRRRKFHKPITMTIPVPPPS
 SEG
 PRD ccchhhhhhhhhhhhhcc
 MEM
 SEQ GEGVSNYGKDTTPNLRLLCSITGGTSPAQWEDITGTTPLTFIKDCVSFTTNVSARFWLA
 SEG
 PRD cc
 MEM
 SEQ DCHQVLETVGLATQLYRELICVPYMAKFVVFAMNDPVESSLRCFCMTDDKVDKTLEQQE
 SEG
 PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhheccccchhhhhhhhhccccchhhhhhhhh
 MEM
 SEQ NFEEVARSKDIEVLEGKPIYVDCYGNLAPLTKGGQQLVFNFYSFKENRPLPFSIKIRDTSQ
 SEG
 PRD cc
 MEM
 SEQ EPCGRSLFLEKPKTTKGLPQTAVCNLNLITLPAHKKIEKTPGRQSFASLALRKRYSYLTP
 SEG
 PRD cc
 MEM
 SEQ GMSPPQSPCERTDIRMAIVADHLGLSWTELARELNFSVDEINQIRVENPNSLISQSFMLK
 SEG
 PRD cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhheccccchhhhhhhhhhhhhhhhh
 MEM


```
SEQ      KVVTRDGKNATTDALTSVLTGINRIDIVTLLEGPIFDYGNISGTRSFADENN VFHDPVDG
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      YPSLQVELETPTGLHYTPPTPFQDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSAD
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

SEQ      GPPVVTAEDASLEDSKLEDSVPLTEMPEAVM
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccc
MEM      .....
```

(No Prosite data available for DKFZphfd2_24p5.3)

(No Pfam data available for DKFZphfd2_24p5.3)

DKFZphfkd2_3113

group: transmembrane protein

DKFZphfkd2_3113 encodes a novel 406 amino acid protein with *C. elegans* cosmid Y37D8A and *A. thaliana* H71412 hypothetical protein.

The novel protein contains 3 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker for kidney cells.

similarity to *A.thaliana* and *C.elegans*;
membrane regions: 3

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="17"

Insert length: 2052 bp

Poly A stretch at pos. 2032, no polyadenylation signal found

```
1 AGTGACGTGA GCGGGTTCGG GTTGTCTGGA GCCCAGCGGC GGGTGTGAGA
51 GTCCGTAAGG AGCAGCTTCC AGGATCCTGA GATCCGGAGC AGCCGGGGTC
101 GGAGCGGCTC CTCAGAGATT ACTGATCTAT GAAATGGCAG AGAATGGAAA
151 AAATTGTGAC CAGAGACGTG TAGCAATGAA CAAGGAACAT CATATGGAA
201 ATTTACACAGA CCCCTCTTCA GTGAATGAAA AGAAGAGGAG GGAGCGGGAA
251 GAAAGGCAGA ATATTGTCTT GTGGAGACAG CCGCTCATTG CCTTGCAGTA
301 TTTTCTCTCT GAAATCCTTG TAATCTTGAA GGAATGGACC TCAAAATTAT
351 GGCATCGTCA AAGCATTGTG GTGTCTTTT TACTGCTGCT TGCTGTGCTT
401 ATAGCTACGT ATTATGTTGA AGGAGTGCAT CAACAGTATG TGCAACGTAT
451 AGAGAAACAG TTTCTTTTGT ATGCCTACTG GATAGGCTTA GGAATTTTGT
501 CTCTCTGTTG GCTTGGAAAC GGGCTGCACA CCTTCTGCTT TTATCTGGGT
551 CCACATATAG CCTCAGTTAC ATTAGCTGCT TATGAATGCA ATTCAGTTAA
601 TTTTCCCGAA CCACCTTATC CTGATCAGAT TATTTGTCCA GATGAAGAGG
651 GCACTGAAGG AACCATTTTT TTGTGGAGTA TCATCTCAAA AGTTAGGATT
701 GAAGCCTGCA TGTGGGGTAT CGGTACAGCA ATCGGAGAGC TGCTCCATA
751 TTTTATGGCC AGAGCAGCTC GCCTCTCAGG TGCTGAACCA GATGATGAAG
801 AGTATCAGGA ATTTGAAGAG ATGCTGGAAC ATGCAGAGTC TGCACAAGAC
851 TTTGCCTCCC GGGCCAAACT GGCAGTTCAG AAACAGTAGC AGAAAGTTGG
901 ATTTTTTGGG ATTTTGGCCT GTGCTTCAAT TCCAAATCCT TTATTTGATC
951 TGGCTGGAAT AACGTGTGGA CACTTTCTGG TACCTTTTGT GACCTTCTTT
1001 GGTGCAACCC TAATTGGAAA AGCAATAATA AAAATGCATA TCCAGAAAAT
1051 TTTTGTATTA ATAACATCA GCAAGCACAT AGTGGAGCAA ATGGTGCTT
1101 TCATTGGTGC TGTCCCCGGC ATAGGTCCAT CTCTGCAGAA GCCATTTTCA
1151 GAGTACCTGG AGGCTCAACG GCAGAAAGCT CACCACAAAA GCGAAATGGG
1201 CACACCACAG GGAGAAAACG GGTGTGCTGT GATGTTTGAA AAGTTGGTCG
1251 TTGTCAATGG GTGTTACTTC ATCCTATCTA TCATTAATCT CATGGCACA
1301 AGTTATGCCA AACGAATCCA GCAGCGGTG AACCTCAGAG AGAAAACATA
1351 ATAAGTAGAG AAAGTTTAA ACTGCAGAAA TTGGAGTGA TGGGTTCTGC
1401 CTTAAATTTG GAGGACTCCA AGCCGGGAAG GAAAATTTCC TTTTCCAACC
1451 TGTATCAATT TTTACAACCT TTTCTCTGAA AGCAGTTTAG TCCATACCTT
1501 GCACTGACAT ACTTTTCTCT TCTGTGCTAA GGTAAGGTAT CCACCTCGA
1551 TGCAATCCAC CTGTGTTTTT CTAGGGTGG AATGTGATGT TCAGCAGCAA
1601 ACTTGCAACA GACTGGCCTT CTGTTTGTTA CTTTCAAAAG GCCCATGA
1651 TACAATTAGA GAATTTCCAC CGCACAAAAA AAGTTCCTAA GTATGTTAAA
1701 TATGTCAAGC TTTTGGGCTT TGTCAAAAT GATTGCTTTG TTTTCTTAAG
1751 TCATCAAAAT GTATATAAAT TATCTAGATT GGATAACAGT CTGTCATGTT
1801 TATCATGTTA CAATTTAATA TTCCATCCTG CCCAACCTCT CCTCTCCCAT
1851 CCTCAAAAAA GGGCCATTTT ATGATGCATT GCACACCCTC TGGGGAAATT
1901 GATCTTTTAA TTTTGAGACA GTATAAGGAA AATCTGGTTG GTGCTTTACA
1951 AGTGAGCTGA CACCATTTTT TATCTGTGTT ATTTAGGATG AAGTCTTGAA
2001 AAAAATCTTA TAAAGACATC TTTAATCATT CCAAAAAAAA AAAAAAAA
2051 AA
```

BLAST Results

Entry AC004686 from database EMBL:

*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 17, clone
hRPC.1073 F 15; HTGS phase 1, 8 unordered pieces.
Score = 4142, P = 6.1e-199, identities = 830/832

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 1351 bp; peptide length: 406
 Category: similarity to unknown protein

```

1 MAENGKNCQ RRVAMNKEHH NGNFTDPSSV NEKKRRERE RQNIVLWRQP
51 LITLQYFSLE ILVILKEWTS KLWHRQSIVV SFLLLAVLI ATYYVEGVHQ
101 QYVQRIEKQF LLYAYWIGLG ILSSVGLGTG LHTFLYLGP HIASVTLAAY
151 ECNSVNFPEP PYPDQIICPD EEGTEGTIFL WSIISKVRIE ACMWGIGTAI
201 GELPPYFMAR AARLSGAEPD DEEYQEFEEF LEHAESAQDF ASRAKLAVQK
251 LVQKVGFFGI LACASIPNPL FDLGITCGH FLVPFWTFFG ATLIGKAIK
301 MHIQKIFVII TFSKHIVEQM VAFIGAVPGI GPSLQKPFQE YLEAQRQKLH
351 HKSEMGTPOG ENWLSWMFEK LVVVMVCYFI LSIINSMAQS YAKRIQQLN
401 SEEKTK

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfd2_3i13, frame 2

TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid
 Y37D8A, N = 1, Score = 905, P = 8.8e-91

TREMBL:ATAC98_2 gene: "YUP8H12.2"; Arabidopsis thaliana chromosome 1
 YAC yUP8H12 complete sequence., N = 1, Score = 470, P = 1.1e-44

PIR:H71412 hypothetical protein - Arabidopsis thaliana, N = 1, Score =
 293, P = 6e-24

>TREMBL:CEY37D8A_20 gene: "Y37D8A.22"; Caenorhabditis elegans cosmid
 Y37D8A
 Length = 457

HSPs:

Score = 905 (135.8 bits), Expect = 8.8e-91, P = 8.8e-91
 Identities = 167/317 (52%), Positives = 228/317 (71%)

```

Query: 38 REERQNIVLWRQPLITLQYFSLEILVILKEWTSKLWHRQSIVVSFLLLLAVLIATYYVEG 97
      R ER+ IV WR+P I + Y +EI + E K+ +++++ + + + + Y+ G
Sbjct: 93 RMERETIVFWRPHIVIPYALMEIAHLAVELFFKILAHKTVLLLTASIGLAVYGYHAPG 152

Query: 98 VHQQYVQRIEKQFLLYAYWIGLGILSSVGLGTGLHTFLLYLGPHIASVTLAAAYECNSVNF 157
      HQ++VQ IEK L +++W+ LG+LSS+GLG+GLHTFL+YLGPHIA+VT+AAAYEC S++F
Sbjct: 153 AHQEHVQTIEKHILWWSWWVLLGVLSSIGLGSGLHTFLIYLGPHIAAVTMAAYECQSLDF 212

Query: 158 PEPPYPDQIICPDDEEGTEGTIFLWSIISKVRIEACMWGIGTAIGELPPYFMAARAARLSGA 217
      P+PPYP+ I CP + + F W I++KVR+E+ +WG GTA+GELPPYFMAARAAR+SG
Sbjct: 213 PQPPYPESIQCPSKSSIAVTF-WQIVAKVRVESLLWGAGTALGELPPYFMAARAARISGQ 271

Query: 218 EPDDEEYQEFEEMLE-HAESAQD----FASRAKLAVQKLVQKVGFFGILACASIPNPLFD 272
      EPDDEEY+EF E++ ES D RAK V+ + ++GF GIL ASIPNPLFD
Sbjct: 272 EPDDEEYREFLELMNADKESDADQKLSIVERAKSWVEHNIHRLGFP GILLFASIPNPLFD 331

Query: 273 LAGITCGHFLVPFWTFFGATLIGKAIKMHQKIFVIITFSKHIVEQMVAFIGAVPGIGP 332
      LAGITCGHFLVPFW+FFGATLIGKA++KMH+Q FVI+ FS H E V + +P +GP
Sbjct: 332 LAGITCGHFLVPFWSFFGATLIGKALVKMHVQMGFVILAFSDHHAENFVKILEKIPAVGP 391

Query: 333 SLQKPFQEYLEAQRQKLH 350
      +++P + LE QR+ LH
Sbjct: 392 YIRQIPISDLLEKQRKALH 409

```

Pedant information for DKFZphfd2_3i13, frame 2

Report for DKFZphfd2_3i13.2

```

SEQ      MAENGKNCQRRVAMNKEHHNGNFTDPSSVNEKKRREREERQNIVLWRQPLITLQYFSLE
SEG      .....xxxxxxxxxxxx
PRD      cccccchhhhhhhhhhhhhcccccceccccchhhhhhhhhhhhhcccccchhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      ILVILKEWTSKLWHRQSI VVSFLLLVAVLIATYVGVHQQYVQRIEKQFLLYAYWIGLG
SEG      .....xxxx
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheecchhhhhhhhhhhhhhhhhhhhhhh
MEM      MM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      ILSVGLGTGLHTFLLYLGPFIASVTLAAEYECNSVNFPEPPYPDQIICPDEEGTEGTIFL
SEG      xxxxxxxxxxxx
PRD      hccccccccceeeeeeccchhhhhhhhhhhhhcccccceccccccccccccccccccccceeee
MEM      .....

SEQ      WSIISKVRIEACMWIGTIGTAIGELPPYFMARAARLSGAEPDDEEYQEFEEMLEHAESAQDF
SEG      .....xxxxxxxxxxxxxxxx
PRD      eehhhhhhhhhhhhhcccccceccccchhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhh
MEM      .....

SEQ      ASRAKLAVQKLQKGVFFGILACASIPNPLFDLAGITCGHFLVPWFTFFGATLIGKAIK
SEG      .....
PRD      hhhhhhhhhhhhhhhccceeeeeeccccccccccccccccceeeeee hhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMM

SEQ      MHIQKIFVITFSKHIVEQMVAFIGAVPGIGPSLQKPFQYEYLAQRQLHHKSEMGTPQG
SEG      .....
PRD      hhhhheeeeeeecchhhhhhhhhhhhhcccccceccccchhhhhhhhhhhhhhhhhhhhhccccc
MEM      .....

SEQ      ENWLSWMFEKLVVVMVCYFILSIINSMACSYAKRIQORLNSEEKTK
SEG      .....
PRD      cchhhhhhhhhhhheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc
MEM      .....

```

PS00001	23->27	ASN_GLYCOSYLATION	PDOC00001
PS00005	69->72	PKC_PHOSPHO_SITE	PDOC00005
PS00006	29->33	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	236->240	CK2_PHOSPHO_SITE	PDOC00006
PS00008	120->126	MYRISTYL	PDOC00008
PS00008	126->132	MYRISTYL	PDOC00008
PS00008	173->179	MYRISTYL	PDOC00008
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	197->203	MYRISTYL	PDOC00008
PS00008	259->265	MYRISTYL	PDOC00008
PS00008	275->281	MYRISTYL	PDOC00008
PS00008	325->331	MYRISTYL	PDOC00008
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	356->362	MYRISTYL	PDOC00008

399

DKFZphfkd2_3o17

group: metabolism

DKFZphfkd2_3o17 encodes a novel 72 amino acid protein with similarity to bos taurus NADH-ubiquinone oxidoreductase B33 subunit (EC 1.6.5.3) (EC 1.6.99.3).

NADH:ubiquinone oxidoreductase is the first enzyme in the respiratory electron transport chain of mitochondria. It is a membrane-bound multi-subunit protein. The bovine heart enzyme contains about 40 different polypeptides. The novel protein is the human orthologue of bovine B22.

The new protein can find application in modulation of the respiratory electron transport chain pathways of mitochondria.

strong similarity to bovine NADH-UBIQUINONE OXIDOREDUCTASE B22 subunit

complete cDNA, complete cds, EST hits,
in frame stop codon at ~274 will be checked
ESTs HS1291620/AA883920 show no stop codon at this side

Sequenced by BMF2

Locus: unknown

Insert length: 693 bp
Poly A stretch at pos. 670, polyadenylation signal at pos. 659

```

1 CAGCAGGCGT GCAGTTTCCC GGCTCTCCGC GCGGCCGGGG AAGGTCAGCG
51 CCGTAATGGC GTTCTTGGCG TCGGGACCCCT ACCTGACCCA TCAGCAAAAG
101 GTGTTGCGGC TTTATAAGCG GCGCTACGC CACCTCGAGT CGTGGTGCGT
151 CCAGAGAGAC AAATACCGAT ACTTTGCTTG TTTGATGAGA GCGCGTTTG
201 AAGAACATAA GAATGAAAAG GATATGGCGA AGGCCACCCA GCTGCTGAAG
251 GAGGCCGAGG AAGAATTCTG GTAACGTCAG CATCCACAGC CATACATCTT
301 CCCTGACTCT CCTGGGGGCA CCTCCTATGA GAGATACGAT TGCTACAAGG
351 TCCAGAAATG GTGCTTAGAT GACTGGCATC CTCTGAGAA GGCAATGTAT
401 CCTGATTACT TTGCCAAGAG AGAACAGTGG AAGAACTGC GGAGGAAAG
451 CTGGGAACGA GAGGTTAAGC AGCTGCAGGA GGAAACGCCA CCTGGTGGTC
501 CTTTAACTGA AGCTTTGCCC CCTGCCCGAA AGGAAGGTGA TTTGCCCCCA
551 CTGTGGTGGT ATATTGTGAC CAGACCCCGG GAGCGGCCCA TGTAGAAAGA
601 GAGAGACCTC ATCTTTCATG CTTGCAAGTG AAATATGTTA CAGAACATGC
651 ACTTGCCCTA ATAAAAATC AGTAAAAAAA AAAAAAAAAA AAA

```

BLAST Results

Entry S28256 from database PIR:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
>TREMBL:MIPTCIB22_1 gene: "ci-B22"; product: "NADH-ubiquinone
oxidoreductase complex B22 subunit"; B.taurus mitochondrion ci-B22
mRNA for B22 subunit of the NADH-ubiquinone oxidoreductase complex
Score = 933, P = 5.2e-93, identities = 163/179, positives = 172/179,
frame +2

Medline entries

92389317
Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from RT bovine heart mitochondria.
Application of a novel strategy for RT sequencing proteins using the polymerase chain reaction

Peptide information for frame 2

ORF from 56 bp to 271 bp; peptide length: 72
Category: strong similarity to known protein

```

1 MAFLASGPYL THQQKVLRLY KRALRHLESW CVQRDKYRYF ACLMRARFEE
51 HKNEKDMAKA TQLLKEAEEE FW*ROHPQPY IFPDSPGGTS YERYDCYKVP
101 EWCLDDWHPS EKAMYPDYFA KREQWKKLRR ESWEREVKQL QEETPPGGPL
151 TEALPPARKE GOLPPLWYI VTRPRERPM

```

BLASTP hits

Sequences producing significant alignments: (bits) Value

sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE... 141 7e-34
tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO NADH-UBIQ... 53 3e-07

>sp|Q02369|NI2M_BOVIN|OD36CE17281FB735 (NDUFB9..)NADH-UBIQUINONE
OXIDOREDUCTASE B22 SUBUNIT (EC 1.6.5.3) (EC 1.6.99.3)
(COMPLEX I-B22) (CI-B22).[BOS TAURUS]
Length = 178

Score = 141 bits (351), Expect = 7e-34
Identities = 63/71 (88%), Positives = 68/71 (95%)

Query: 2 AFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKAT 61
AFL+SG YLTHQQKVLRLYKRALRHLESWC+ RDKYRYFACL+RARF+EHKNEKDM KAT
Sbjct: 1 AFLSSGAYLTHQQKVLRLYKRALRHLESWCIRDKYRYFACLLRARFDEHKNEKDMVKAT 60

Query: 62 QLLKEAEEEEFW 72
QLL+EAEFEFW
Sbjct: 61 QLLREAEEEEFW 71

>tr|U41534|Q18036|D34BCCB6E8FBCD5F (C16A3.4)SIMILAR TO
NADH-UBIQUINONE OXIDOREDUCTASE B22.[CAENORHABDITIS
ELEGANS]
Length = 163

Score = 52.7 bits (124), Expect = 3e-07
Identities = 25/64 (39%), Positives = 41/64 (64%), Gaps = 1/64 (1%)

Query: 10 LTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKATQLLKEAE 68
L+H+QKV RLYKR LR +++W + + R+ C++RARF+ + +E D K+ LL +
Sbjct: 12 LSHRQKVTRLYKRCLREVDNWWYGGNNLEVRVQKCIIRARFDANAEVDTRKSQILLADGC 71

Query: 69 EEFW 72
+ W
Sbjct: 72 RQLW 75

Alert BLASTP hits for DKFZphfkd2_3o17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_3o17, frame 2

Report for DKFZphfkd2_3o17.2

[LENGTH] 72
[MW] 8839.28
[pI] 9.26
[HOMOL] PIR:S28256 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain CI-B22 - bovine
2e-34
[KW] All_Alpha

SEQ MAFLASGPYLTHQQKVLRLYKRALRHLESWCVQRDKYRYFACLMRARFEEHKNEKDMAKA
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhh

SEQ TQLLKEAEEEEFW
PRD hhhhhhhhhccc

(No Prosite data available for DKFZphfkd2_3o17.2)

(No Pfam data available for DKFZphfkd2_3o17.2)

DKFZphfkd2_46a6

group: kidney derived

DKFZphfkd2_46a6 encodes a novel 315 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="228.6 cR from top of Chr15 linkage group"

Insert length: 2774 bp

Poly A stretch at pos. 2751, polyadenylation signal at pos. 2732

```
1 CTGCGGAGCG CAGCTATGGC TGCTGGCGTA CCCTGTGCGT TAGTCACCAG
51 CTGCTCCTCC GTCTTCTCAG GAGACCAGCT GGTCCAACAT ACCCTTGGAA
101 CAGAAGATCT TATTGTGGAA GTGACTTCCA ATGATGCTGT GAGATTTTAT
151 CCCTGGACCA TTGATAATAA ATACTATTCA GCAGACATCA ATCTATGTGT
201 GGTGCCAAAC AAATTTCTTG TTAGTCGAGA GATTGCAGAA TCTGTCCAAG
251 CATTTGTGGT TTAATTTGAC AGCACACGAA AATCGGGCCT TGATAGTGTC
301 TCCTCATGGC TTCCACTGGC AAAAGCATGG TTACCTGAGG TGATGATCTT
351 GGTCTGCGAT AGAGTGTCTG AAGATGGTAT AAACCGACAA AAAGCTCAAG
401 AATGGAGCCT CAAACATGGC TTGAATTGG TAGAAGTTAG TCCAGAGGAG
451 TTGCTGAGG AGGATGATGA CTTCACAGAA TCTACAGGAG TAAAGCGAAT
501 TGTCCAAGCC CTGAATGCCA ATGTGTGGTC CAATGTAGTG ATGAAGAATG
551 ATAGGAACCA AGGCTTTAGC CTCTCAACT CATTGACTGG AACAAACCAT
601 AGCATTGGGT CAGCAGATCC CTGTCAACCA GAGCAACCCC ATTTGCCAGC
651 AGCAGAGTAG ACTGAATCCC TCTCTGATCA TCGGGGTGGT GCATCTAACA
701 CAACAGATGC CCAGGTTGAT AGCATTGTGG ATCCCATGTT AGATCTGGAT
751 ATTCAGGAAT TAGCCAGTCT TACCACTGGA GGAGGAGATG TGGAGAATTT
801 TGAAGAGCCC TTTTCAAGT TAAAGGAAAT GAAAGACAAG GCTGCGACGC
851 TTCCTCATGA GCAAAGAAAA GTGCATGCAG AAAAGTGGC CAAAGCATTC
901 TGGATGGCAA TCGGGGGAGA CAGAGATGAA ATTGAAGGCC TTTCATCTGA
951 TGGAGAGCAC TGAATTATTC ATACTAGGGT TTGACCAACA AAGATGCTAG
1001 CTGTCTCTGA GATACCTCTC TACTCAGCCC AGTCATATTT TGCCAAAATT
1051 GCCCTTATCA TGTGTGGCTG CTGACTTGTT TATAGGGTCC CCTTAATTTT
1101 AGTTTTTTAGT AGGAGGTTAA GGAGAAATCT TTTTTCCT CAGTATATTG
1151 TAAGAGAGTG AGGAATACAG TGATAGTAAT GAGTGAGGAT TTCTTAAATA
1201 TACTTTTTTT TTGTTCTAGG AATGAGGGTA GGATAAATCT CAGAGGCTGT
1251 TGTGATTTAC TCAAGTTGAA GACAACCTCC AGGCCATTCC TGGTCAACCT
1301 TTTAAGTAGC ATTTCCAGCA TTCACACTTG ATACTGCACA TCAGGAGTTG
1351 TGTCACTTTT CCTGGGTGAT TTGGGTTTTT TCCATTCAAG GAGCTTGAG
1401 CTCTGAGCTA TGATGCTTTT ATTGGGAGGA AAGGAGGCAG CTGCAGAATT
1451 ATGTGTGAGCT ATGTGGGGCC GAAGTCTCAG CCCGCAGCTA AGTCTCTACC
1501 TAAGAAATG CCTCTGGGCA TTCTTTTGAA GTATAGTGTC TGAGCTCATG
1551 CTAGAAAGAA TCAAAAAGCC AGTGTGGATT TTAGGCTGT AATAAATGAG
1601 GCAAAGGATT TCTATTCCAG TGGGAAGGAA ACCTCTCTAC TGAGTTGTGG
1651 GGGATATGTT GTATGTTAGA GAGAACCCTA AGGAGTCCCT GTATGGGCCA
1701 TGGAGACAGT ATGTGATAAC ATACCGTGAT TTTATGAAG AAATTTCTCT
1751 GTCCTAGAGT TCTCCCCGTC TGCTTGAGAT GCCAGAGCTG TGTTGTTGCA
1801 CACCTGCAAA ACAAGGCACA TTTCCCCCTT TCTCTTAAAG GCCAAGAGAG
1851 GATCACTGCC AAAGTGGGAG CACTAAGGGG TGGGTGGGGA AGTGAAATGT
1901 TAGGCGATGA ATTCTGAGC ACCTTGTTTT TCTTCCAAGG TTCGTAGCTC
1951 CTCTCTGCCC TTCCAAGCCT GTAACCTCGG AGGACTATCT TTTGTTCTCT
2001 ATCCTTTGTC TTGTTAGAGT GGGTCAGCCC CAGAGGAACT GATAAGCAAA
2051 TGGCAAGTTT TTAAGGAAG AGTGGAAAGT ACTGCAATA AAAATCCTTA
2101 TTTGTTTTTG TAGACTTTGT AATGCATATC ATTAGCCCTC ACTGTGATCA
2151 TTAAGTCTGT GGCTCTGAAC TGCCACATAG TACAGTGGAT GGAAGGTGCC
2201 CGCACACCAG CTGAGAACTG GTTCTGGCCT AGGTGGGCTC TAGAACCATT
2251 TACACAGCAT GAAAGAAACA GGTGGGTTA GGAGCAGAAA GAAATAAGGC
2301 TCACACCCCT CCAGACACTA CTTATAAGC ACTGCAGAAC CTGAAACAGA
2351 TGGCAGAAAG AATGGAATGC TACAGGGGCC AGCAGGAGTG ACCACAGGGA
2401 GGGGACAGCT CAGTGACTGG AGCATTGAGG AAGAGGCTTT CCAGGGAACA
2451 CTGGACATTG CTTAGTGACC TTTTGTTCTT TTTTCTTTT TTTTCTTTT
2501 CTGTTCTGAA AGACTTTGAG TCTGTGGTTC ACCACGAGCC CATCAGTGTT
2551 TCTTTGAGGT GATTGCATTA GGAAGTTGG CTCTGGGATT GCAAAAAAAA
2601 AAAAAAGGTG GAACATGTTT TCCTTAAAG ATGGAAGGTT TTAGAAAAA
2651 TACTAGGCCA TCTGGTTAGA AAAACAGAC CAGACTAGAA AAAGCTGTGA
```

2701 ATTTGATTTT GTAGATTAAA CAAAGCCAGA TGATTAAAT GTGATTATT
2751 TATAAAAAAA AAAAAAAAAA AAAA

BLAST Results

Entry HS463358 from database EMBL:
human STS WI-14364.
Length = 472
Minus Strand HSPs:
Score = 1605 (240.8 bits), Expect = 5.0e-68, P = 5.0e-68
Identities = 347/361 (96%)

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 16 bp to 960 bp; peptide length: 315
Category: putative protein
Classification: unset

1 MAAGVPCALV TSCSSVFSGD QLVQHTLGE DLIVEVTSND AVRFYPWTID
51 NKYYADINL CVVPNKFLVT AEIAESVQAF VVYFDSTRKS GLDSVSSWLP
101 LAKAWLPEVM ILVCDRVSED GINRQKAQEW SLKHGFELVE LSPEELPEED
151 DDFPESTGVK RIVQALNANV WSNVVMKNDR NQGFSLNLSL TGTNHSIGSA
201 DPCHPPEQPHL PAADSTESLS DHRGGASNTT DAQVDSIVDP MLDLDIQELA
251 SLTTGGGDVE NFERPFSLKL EMKDKAATLP HEQRKVHAEK VAKAFWMAIG
301 GDRDEIEGLS SDGEH

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46a6, frame 1

PIR:T04362 probable GTP-binding protein yptm3 - maize, N = 1, Score =
87, P = 0.21

PIR:S71585 GTP-binding protein GB2 - Arabidopsis thaliana, N = 1, Score
= 86, P = 0.27

>PIR:T04362 probable GTP-binding protein yptm3 - maize
Length = 210

HSPs:

Score = 87 (13.1 bits), Expect = 2.4e-01, P = 2.1e-01
Identities = 34/160 (21%), Positives = 67/160 (41%)

Query: 48 TIDNKYYADINLCVVPNKFL-VTAEIAESVQAFVVYFDSTRKSGLDSVSSWLP LAKAWL 106
TIDNK I F +T ++ +D TR+ ++SWL A+
Sbjct: 49 TIDNKP I K L Q I W D T A G Q E S F R S I T R S Y Y R G A A G A L L V Y D I T R R E T F N H L A S W L E D A R Q H A 108
Query: 107 PE---VMIL--VCDRVSEDGINRQKAQEWSLKHGFELVELSPEELPEEDDDFFPESTGVKR 161
VM++ CD ++ ++ +++ +HG +E S + ++ F ++ G
Sbjct: 109 NANMTVM L I G N K C D L S H R R A V S Y E E G E Q F A K E H G L V F M E A S A K T A Q N V E E A F I K T A G T -- 166
Query: 162 IVQALNANVWSNVVMKNDRNQGFSLNLSLTGTNHSIGSADPC 203
I + + ++ N G+++ NS G S A C
Sbjct: 167 IYKKIQDGIFDVS NESNGIKVGYAVPNSSGGGAGSSSQAGGC 208

Pedant information for DKFZphfkd2_46a6, frame 1

Report for DKFZphfkd2_46a6.1

[LENGTH] 315

[MW] 34505.54
[pI] 4.55
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.67 %

SEQ MAAGVPCALVTSCSSVFSGDQLVQHTLGTEDLIVEVTSNDAVRFPWTIDNKYYADINL
SEG
PRD cccccceeeccccccccceeeccccceeeccccceeeccccccccccccce

SEQ CVVPNKFLVTAEIAESVQAFVVYFDSTRKSGLDVSSWLPKAWLPPEVMILVCDRVSED
SEG
PRD eeccccchhhhhhhhhheeeccccccccccccccccccccceeeccccccc

SEQ GINRQKAQEWSLKHGFELVELSPEELPEEDDFPESTGVKRIVQALNANVSNVVMKNOR
SEGxxxxxxxxxxxxxxxxxxxxx.....
PRD cchhhhhhhhhccccceeeccccccccccccccccchhhhhhhccccceeecccc

SEQ NQGFSLLNSLTGTNHSIGSADPCHPEQPHLPAADSTESLSDHRGGASNTTDAQVDSIVDP
SEG
PRD cch

SEQ MLDLDIQELASLTGGGDVENFERPFPSKLEMKDKAATLPHEQRKVHAEKVAKAFWMAIG
SEG
PRD hhhhhhhhhhhccccccccccccchhhhhhhhhhhhhccchhhhhhhhhhhhhhhc

SEQ GDRDEIEGLSSDGEH
SEG
PRD ccccccccccccc

(No Prosite data available for DKFZphkd2_46a6.1)

(No Pfam data available for DKFZphkd2_46a6.1)

DKF2phfkd2_46b10

group: kidney derived

DKF2phfkd2_46b10.1 encodes a novel 315 amino acid protein with similarity to C.elegans cosmid F25B5.3

The novel protein contains a HTH-LYSR-family PROSITE pattern. Proteins of the lysR family are bacterial transcriptional regulatory proteins which bind DNA using a helix-turn-helix motif. Most of these proteins are transcription activators and usually negatively regulate their own expression. They all possess a potential 'helix-turn-helix' DNA-binding motif in their N-terminal section. The 'helix-turn-helix' motif is missing in DKF2phfkd2_46a6.1. No informative BLAST results, no predictive PFAM or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to C.elegans F25B5.3

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 1285 bp

Poly A stretch at pos. 1266, no polyadenylation signal found

```
1 CAGTCTACGC GAGCTGCCTG TTTTTCCT GCTGGACGC GCATGAGGGC
51 CCCGTCCATG GACCGCGCGG CCGTGGCGAG GGTGGCGCG GTAGCGAGCG
101 CCAGCGTGTG CGCCCTGGTG GCGGGGGTGG TGCTGGCTCA GTACATATTC
151 ACCTTGAAGA GGAAGACGGG GCGGAAGACC AAGATCATCG AGATGATGCC
201 AGAATTCCAG AAAAGTTCAG TTCGAATCAA GAACCTTACA AGAGTAGAAG
251 AAATTATCTG TGGTCTTATC AAAGGAGGAG CTGCCAACT TCAGATAATA
301 ACGGACTTTG ATATGACACT CAGTAGATTT TCATATAAAG GGAAGAGATG
351 CCCAACATGT CATAATATCA TTGACAACTG TAAGCTGGTT ACGGATGAAT
401 GTAGAAAAAA GTTATTGCAA CTAAGGAAA AATATTACGC TATTGAAGTT
451 GATCCTGTTC TTACTGTAGA AGAGAAGTAC CCTTATATGG TGGATGGTA
501 TACTAAATCA CATGGTTTGC TTGTCAGCA AGCTTTACCA AAAGCTAAAC
551 TTAAGAAAT TGTGGCAGAA TCTGACGTTA TGCTCAAAGA AGGATATGAG
601 AATTTCTTTG ATAAGCTCCA ACAACATAGC ATCCCGTGT TCATATTTTC
651 GCGTGGAAAT GCGGATGTAC TAGAGGAAGT TATTCGTCAA GCTGGTGTTC
701 ATCATCCCAA TGTCAAAGTT GTGTCCAATT TTATGGATTT TGATGAACT
751 GGGGTGCTCA AAGGATTTAA AGGAGAACTA ATTCATGTAT TTAACAAACA
801 TGATGGTGCC TTGAGGAATA CAGAATATTT CAATCAACTA AAAGACAATA
851 GTAAACATAAT TCTTCTGGGA GACTCCCAAG GAGACTTAAG AATGGCAGAT
901 GGAGTGGCCA ATGTTGAGCA CATTCTGAAA ATTGGATATC TAAATGATAG
951 AGTGGATGAG CTTTGTAGAA AGTACATGGA CTCTTATGAT ATTGTTTTAG
1001 TACAAGATGA ATCATTAGAA GTAGCCAACT CTATTTTACA GAAGATTCTA
1051 TAAACAAGCA TTCTCCAAGA AGACCTCTCT CCTGTGGGTG CAATTGAACT
1101 GTTCATCCGT TCATCTTGTG GAGAGACTTA TTTTAAATAT ATCCTTACTC
1151 TCGAAGTGT CCCTTTGTAT AACTGAAGTA TTTTCAGATA TGGTGAATGC
1201 ATTGACTGGA AGCTCCTTTT CTCCACCTCT CTCACACAC TCCTCACCGT
1251 ATCTTTTAA CCATTTAAAA AAAAAAAAAA AAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 43 bp to 1050 bp; peptide length: 336

Category: similarity to unknown protein

Classification: unset

Prosite motifs: HTH_LYSR_FAMILY (16-47)

```
1 MRAPSMDDRAA VARVGAVASA SVCALVAGVV LAQYIFTLKR KTGKTKIIE
51 MMPEFQKSSV RIKNPTRVEE IICGLIKGGA AKLQIITDFD MTLRSFSYKG
101 KRCPTCHNII DNCKLVTDEC RKKLLQLKEK YYAIEVDPVL TVEEKYPYV
151 EWYTKSHGLL VQQALPKAKL KEIVAESDVM LKEGYENFFD KLQQHSIPVF
201 IFSAGIGDVL EEVIRQAGVY HPNVKVVSNF MDFDETGVLK GFKGELIHVF
251 NKHDGALRNT EYFNQLKDNS NIILLGDSQG DLRMADGVAN VEHLKIGYL
301 NDRVDELLEK YMDSYDIVLV QDESLEVANS ILQKIL
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phkd2_46b10, frame 1

SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III., N = 1, Score = 524, P = 2.2e-50

TREMBL:AC005499_12 gene: "T6A23.12"; Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence, complete sequence., N = 2, Score = 194, P = 1.4e-26

>SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.

Length = 376

HSPs:

Score = 524 (78.6 bits), Expect = 2.2e-50, P = 2.2e-50
Identities = 112/300 (37%), Positives = 174/300 (58%)

```
Query: 44 RKTIIEMMPEFQ--KSSVRIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYK-G 100
+KT ++ ++ + + + + +PT V + ++ GGA K +I+DFD TLSRF+ + G
Sbjct: 73 KKTDOVPLLMNYLLGEEQILVADPTAAAKLRKMVVGAGAGT VVISDFDYTLRSFANEQG 132

Query: 101 KRCPTCHNIID-NCKLVTDEC RKKLLQLKEKYYAIEVDPVL TVEEKYPYVWEYTKSHGL 159
+R T H + D N + E +K + LK KYY IE P LT+EEK P+M +W+ SH L
Sbjct: 133 ERLSTTHGVFDDNVMLRKLPELGQKFVDLKNKYPIEFSPNLTMEKIPHMEKWWGTSHSL 192

Query: 160 LVQQALPKAKLKEIVAESDVM LKEGYENFFD KLQQHSIPVFI FSAGIGDVL EEVIRQA-G 218
+V + K +++ V +S ++ K+G E+F + L H+IP+ IFSAGIG+++E ++Q G
Sbjct: 193 IVNEFKSKNTIEDFVRQSRIVFKDGAEDFIEALDAHNIPLVIFSAGIGNIIEYFLQKQLG 252

Query: 219 VYHPNVKVVSNFMDFDETGV LKGFKGELIHVFNKHDGAL-RNTEYFNQLKDNSNIILLGD 277
N +SN + FDE F LIH F K+ + + T +F+ + N+ILLGD
Sbjct: 253 AIPRNTHFISNMILFDEDDNACAFSEPLIHTFCKNSSVIQKETSFFHDITAGRVNVILLGD 312

Query: 278 SQGDLRMADGVANVEHLKIGYLNDRVDEL--LEKYMDSYDIVLVQDESLEVANSILQKI 335
S GD+ M GV LK+GY N +D+ L+ Y + YDIVL+ D +L VA I+ I
Sbjct: 313 SMGDIHMDVGVVERDGPTLVKGYNGSLDDTAALQHYEEVYDIVLIHDPTLNVAQKIVDII 372
```

Pedant information for DKF2phkd2_46b10, frame 1

Report for DKF2phkd2_46b10.1

```
[LENGTH] 336
[MW] 37948.37
[PI] 6.67
[HOMOL] SWISSPROT:YQT3_CAEEL HYPOTHETICAL 42.0 KD PROTEIN F25B5.3 IN CHROMOSOME III.
3e-51
[PROSITE] HTH_LYSR_FAMILY 1
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 7.44 %
```

```
SEQ MRAPSMDDRAAVARVGAVASASVCALVAGVVLAQYIFTLKRKTGRKTKIEMMPEFQKSSV
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccchhhhhcchhhhhhhheehhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceehhhhhhhhhhe
MEM .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ RIKNPTRVEEIIICGLIKGGA AKLQIITDFD MTLRSFSYKGRKRCPTCHNIIDNCKLVTDEC
SEG .....
PRD eccccchhhhhhhhhcccccceeeccccccccccccccccccccccccccccccccchhhhh
MEM .....cccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```

SEQ      RKKLLQLKEKYAIEVDPVLTVEEKYPYMWVYTKSHGLLVQQALPKAKLKEIVAESDVM
SEG      .....
PRD      hhhhhhhhhhhheeeccccccccccchhhhhhhcccccchhhhhccchhhhhhhhhhhcc
MEM      .....

SEQ      LKEGYENFFDKLQQHSIPVFI FSAGIGDVLEEVI RQAGVYHPNVKVVS NFMD FDETGV LK
SEG      .....
PRD      cccccchhhhhhhhhccceeeccccchhhhhhhhhcccccceeeccccccccce
MEM      .....MMMMMMMMMMMMMMMM.....t.....

SEQ      GFKGELIHVFNKHDGALRNTEYFNQLKDNSNILLGDSQGD LRMADGVANVEHILKIGYL
SEG      .....
PRD      eccccceeeccccccccccchhhhhhhceeeccccccccccccccccccccceec
MEM      .....

SEQ      NDRVDELLEKYMDSYDIVLVQDESLEVANSILQKIL
SEG      .....
PRD      cchhhhhhhhhhhheeeecchhhhhhhhhcc
MEM      .....

```

Prosites for DKFZphfkd2_46b10.1

PS00044 16->47 HTH_LYSR_FAMILY PDOC00043

(No Pfam data available for DKFZphfkd2_46b10.1)

DKF2phfkd2_46d13

group: kidney derived

DKF2phfkd2_46d13 encodes a novel 506 amino acid protein with weak similarity to KE03 protein

The novel protein contains a RGD site.

No informative BLAST results; No predictive prosite, pfam or SCOP motive

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KE03 protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="227.6 cR from top of Chr1 linkage group"

Insert length: 3346 bp

Poly A stretch at pos. 3328, polyadenylation signal at pos. 3308

```

1 CTCTCGCGAG AGGAGCAAGA GGAAGATGGC CGTGCCCTGT TTTTCGGTGT
51 AAGGCAGCAG ACGGCGGCTG CGACGGCGAG ACTGAGATCC TGGTGTCTGT
101 GGCACCTGAG TTCTAGCTTC CCCGAGCGAG CGCGCGTCCC TTCTGTCCCTA
151 GCGGAGAGCC GGCTCTTCCC CGGGAGATGC GTTTGTCCCA GGCTCGGGGG
201 CTCAGTGGGA GTTCATGCTG CGCTGGAGGC TCTTGGCCAC CGCTCTAATC
251 GCCTTGTGCC GCCGCAGCGC CAGCTCCGTC GCCAGCGGTG AGCCTCCCGA
301 TTCCCCCCTT TGCCCTGGG GCGCGCGATG ACCGGGGAGA AGATCCGCTC
351 ACTGCGGAGG GACCACAAGC CCAGCAAAGA AGAAGGGGAC CTGCTGGAGC
401 CCGGGGATGA AGAAGCGGCG GCTGCCCTCG CGGTACCTT TACCAGAAGC
451 AGGATTGGCA AGGGCGGCAA AGCTTGTCTA AAGATCTTCA GTAAACATCA
501 CCACCGGCTA CAGCTGAAGG CAGCTCCGCG CTCTCCCAAT CCCCCCGCG
551 CCGCGGCTCT GCCGCTGCAC AATTCCTCCG TGACTGCCAA CTCCAGTCC
601 CCGGCCCTTC TGGCCGGCAC CAACCCCGTT GCTGTCGTCG CGGATGGAGG
651 CAGTTGCCCC GCACACTACC CGGTGCACGA GTGCGTCTTC AAGGGGGATG
701 TGAGGAGACT CTCCTCTCTC ATCCGCACGC ACAATATCGG GCAGAAAGAT
751 AATCAGGAA ATACTCCTTT ACACCTTGCT GTGATGTTAG GAAATAAAGT
801 PACAGCTCTT TTGAGGAAGC TTAAGCAGCA ATCCAGGGAA AGTGTGAAG
851 AAAACGACC TCGATTATTA AAAGCCCTGA AAGAGCTAGG TGACTTTTAT
901 CTAGCACTTC ACTGGGATT TCAAAGCTGG GTGCCTTAC TTTCCGAAT
951 TCTGCCTCC GATGCATGTA AAATATACAA ACAAGGTATC AATATCAGGC
1001 TTGACACAAC TCTCATAGAC TTTACTGACA TGAAGTGCCA ACGAGGGGAT
1051 CTAAGCTTCA TTTCAATGG GCATGCGCGC CCCTCTGAAT CTTTGTAGT
1101 ATTAGCAAT GAACAAAAAG TTTATCAGCG AATACATCAT GAGGAATCAG
1151 AGATGGAAAC AGAAGAAGAG GTGGATATTT TAATGAGCAG TGATATTTAC
1201 TCTGCAACTT TATCAACAAA ATCAATTTCT TTACGCGGTG CCCAGACAGG
1251 ATGGCTTTTT CGGGAAGATA AAACAGAAAG AGTAGGAAAC TTTTGGCAG
1301 ACTTTTACCT GGTGAATGGA CTTGTTATAG AATCAAGGAA AAGAAGAGAA
1351 CATCTCAGTG AAGAGGATAT TCTTCGAAAT AAGGCCATCA TGGAGAGTTT
1401 GAGTAAAGGT GGAAACATAA TGGAACAGAA TTTTGAGCCG ATTCGAAGAC
1451 AGTCTCTTAC ACCGCTCTCT CAGAACACTA TTACATGGGA AGAATATATA
1501 TCTGCTGAAA ATGGAAAAGC TCCTCATCTG GGTAGAGAAAT TGGTGTGCAA
1551 AGAGAGTAAG AAAACGTTTA AAGCTACGAT AGCCATGAGC CAGGAATTTC
1601 CCTTAGGGAT AGAGTTATTA TTGAATGTTT TAGAAGTAGT AGCTCCCTTC
1651 AAGCACTTTA ACAAGCTTAG AGAATTGTGT CAGATGAAGC TTCTCCAGG
1701 CTTTCTGTGA AAATTAGATA TACCTGTGTT TCCCACAACT ACAGCCACTG
1751 TGACTTTTCA GGAGTTTCCA TACGATGAAT TTGATGGCTC CATCTTTACT
1801 ATACCTGATG ACTACAAGGA AGACCCAAGC CGTTTCTCTG ATCTTTAACT
1851 GACGTGGAAA AGGATGCCGT CTAACCAAGG AAAGAAAATA CAGAGACCCCT
1901 AGAAGTGGAT CCAAATAGAA GGGACAAATG CTTTCAGTGA AGAAAAGGGA
1951 ATTACACATT GAATCGACAC ATCAGTAATA CGATACAGTG AAATGGGCCT
2001 CTAATAAGAA TTTACGCGAG TTTTCTGATG TGCCATTTTT TGCTTTTTTA
2051 AAAATATACA TATTATAAAT GTAATAGTTT GACACATTA TGACCCTAAG
2101 ACCTGCGTAT GTGAAGCAGC TATGAGTGCT GTGATTTGTT TTTAAAAATT
2151 TTTACACTTC TTGTTGAAAT ATATATGCAT ATAAATATAT CTATATCTAT
2201 ATCTATATCT AAAACACTCC TGGACCATT ACGTAAATTA AATGTCTTAA
2251 GAGATATGGA GCCCTTTTAA ACTTGTCACT TTTATGCAAG GTGACATTTA
2301 TAAATATTCC TTCGAGCTTT GTTTTCATAA AATGTAAACT ATGTAACATT
2351 ATGTATAGTT CAGTAATTTG AATGTTTGT CAATATAATG AACTAGAAGG
2401 AATGCAATTT TCTGTAGATG AATGAACCAA ATGGTAACCA TTAACAATTT
2451 GCATTTATAT GTTGAATATC ATTTCAAGAG GAGCGTTCAC TCTGCAGGGA
2501 ATAAGGTACC TCCTTTAGCA CCTTAGTGCA ATTCAATTGT GTGCTATTGT
2551 TTTTACCTG AATGTTTGT ACTAATCTTC CTTTCATAGA ACCTCTATTT
2601 TTTTTTTTTC TAAACTTGAG TTTGAGTCTT TGTATATGTC ATCATAAGGT

```

```

2651 AATGGTTAGC ATGTTTAAAG ATATTCTCTC TCCAAATCTC AGCACTTTAA
2701 AAAAAAATCC AAATTTTAA ACTTGCTTCC TAATAAGTAC ACATCGGTCT
2751 GATTATTTTG TTTGTTTTTA GTAGAATATG GATGCATTGG TGTCAGTTTT
2801 AAAAAACAAT ACACATATTT TGGACAACCC TACATATTTA ATCCTTTCAA
2851 AATAAGATAA AAACATTTTA TATGCTAACA GAATATATTT GTTACAAGTT
2901 AAGTCCAGAA AGTATACACA AGATTGATTA CTCTTATTAT TTTTTTAA
2951 TCACAGGAAA ATATTGATTT CATTGCTTCC AAAGTGATAA AATCTTGAT
3001 TACTCATTTT TGCACTTAAA ATTTTCTTA TTTATCCAA GGTGGTTTGA
3051 AGGTCCAAGT ATGAAAATAA ATTAGGGGGA TTAATGTATA ACAGTTATA
3101 AGTATCATGT TGTATTAAAG AGCTTACTTA GATTGATGTT TTTAAATGT
3151 ATCTCATGTA ATGCTCAAG AATGCATCTG TCAAGTTTTT TAGACTGACC
3201 AGTAGCTTAA ACTTTTTTCA GGATTTTAGG TAATTGAAA GGAGTTAGA
3251 GACCTTTATT GAAAATATGA TTTAAATATC CAAAGCATAA ACCGTAAGAA
3301 AAATTTTAAA TAAACATCTT TAAAGCTGAA AAAAAA AAAA

```

BLAST Results

Entry HS121353 from database EMBL:
human STS WI-14729.
Score = 1697, P = 1.9e-69, identities = 363/379

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 328 bp to 1845 bp; peptide length: 506
Category: similarity to unknown protein

```

1 MTGKIRSLR RDHKPSKEEG DLLEPGDEEA AAALGGTFTR SRIGKGGKAC
51 HKIFSNHHHR LQLKAAPASS NPPGAPALPL HNSSVTANSQ SPALLAGTNP
101 VAVVADGGSC PAHYPVHECV FKGDVRRLLSS LIRTHNIGQK DNHGNTPLHL
151 AVMLGNKVTA LLRKLKQQR ESVEEKRPRL LKALKELGDF YLELHWFQFS
201 WVPLLSRILP SDACKIYKQG INIRLDTTLI DFTDMKCQRG DLSFIFNGDA
251 APSESFVVLD NEQKVYQRIH HEESEMETEE EVDILMSSDI YSATLSTKSI
301 SFTRAQTGWL FREDKTERVG NFLADFYLVN GLVIESRKRRL EHLSEEDILR
351 NKAIMESLSK GGNIMEQNFE PIRRQSLTTP PONTITWEEY ISAENGKAPH
401 LGRELVCRES KKTFKATIAM SQEFLPLGIEL LLNVLEVVP FKHFNKLREF
451 VQMKLPPGFP VKLDIPVFPT ITATVTFQEF RYDEFDGSIF TIPDDYKEDP
501 SRFPDL

```

BLASTP hits

Entry CEC01F1_3 from database TREMBL:
gene: "C01F1.6"; Caenorhabditis elegans cosmid C01F1.
Score = 371, P = 4.5e-61, identities = 69/138, positives = 96/138

Entry CEC18F10_9 from database TREMBL:
gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10.
Score = 383, P = 3.4e-39, identities = 103/349, positives = 182/349

Entry AF064604_1 from database TREMBL:
product: "KE03 protein"; Homo sapiens KE03 protein mRNA, partial cds.
Score = 348, P = 8.3e-32, identities = 95/295, positives = 148/295

Alert BLASTP hits for DKFzphfkd2_46d13, frame 1

No Alert BLASTP hits found

Pedant information for DKFzphfkd2_46d13, frame 1

Report for DKFzphfkd2_46d13.1

```

[LENGTH] 506
[MW] 57003.12
[pI] 6.40

```

```

(HOMOL)      TREMBL:CEC18F10_9 gene: "C18F10.7"; Caenorhabditis elegans cosmid C18F10. 2e-35

(BLOCKS)     BL01288E
(PROSITE)    RGD      1
(PROSITE)    MYRISTYL  7
(PROSITE)    CAMP_PHOSPHO_SITE  2
(PROSITE)    CK2_PHOSPHO_SITE   9
(PROSITE)    PKC_PHOSPHO_SITE   6
(PROSITE)    ASN_GLYCOSYLATION  1
(KW)         Alpha_Beta
(KW)         LOW_COMPLEXITY      7.51 %

SEQ  MTGEKIRSLRRDHKPSKEEGDLLEPGDEEAAAALGGTFTSRISRGKGGKACHKIFSNHHHR
SEG  .....XXXXXXXXXXXXX.....
PRD  cccceeeccccccccccccccccchhhhhccccccccccccceeeeeeccchhhh

SEQ  LQLKAAPASSNPPGAPALPLHNSVTANSQSPALLAGTNFVAVVADGGSCPAHYVPVHECV
SEG  ...XXXXXXXXXXXXXXXXX.....
PRD  hhhhhccccccccceeeccccccccccccceeeccccceeeccccccccceee

SEQ  FKGDVRRLLSLIRTHNIGQKDNHGNTPLHLAVMLGNKVTALLRKLKQOSRESVEEKRPRL
SEG  .....
PRD  ecccchhhhhhhhhccccccccccccceeeccccchhhhhhhhhhhccccchhhhhhhhh

SEQ  LKALKELGDFYLELHWDQSWVPLLSRILPSDACKIYKQGINIRLDTTLIDFTDMKQCRG
SEG  .....
PRD  hhhhhccccceeehhhhccccceeeccccccccceeeccccceeecccccccccccc

SEQ  DLSFIFNGDAAPSESFVLDNEQKVYQRIHHEESEMETEEVLDLMSSDIYSATLSTKSI
SEG  .....XXXXXXXXXXXXX.....
PRD  ceeeeccccceeeeeeccccceeehhhhhhhhhhhhhhhhhhhhccccceeecccccc

SEQ  SFTRAQTGWLFREDKTERVGNFLADFYLVNGLVIESKRREHLSEEDILRNKAIMESLSK
SEG  .....
PRD  eeeeeccccccccchhhhhheeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhc

SEQ  GGNIMEQNFEPIRRQLTPPPQNTITWEEYISAENGKAPHLGRELVCESKKTFRKTIAM
SEG  .....
PRD  cceeeccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  SQEFLGIGIELLNVLEVAPFKHFNKLREFVQMKLPFGFPVKLDIPVFPTITATVTQFQF
SEG  .....
PRD  hhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeeeeehhhhhhhcc

SEQ  RYDEFDGSIFTIPDDYKEDPSRFPDL
SEG  .....
PRD  cccccccccccccccccccccccccc

```

Prosites for DKFZphfd2_46d13.1

PS00001	82->86	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	373->377	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	8->11	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	336->339	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	228->232	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	278->282	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	386->390	CK2_PHOSPHO_SITE	PDOC00006
PS00006	476->480	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	46->52	MYRISTYL	PDOC00008
PS00008	108->114	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	487->493	MYRISTYL	PDOC00008
PS00016	239->242	RGD	PDOC00016

(No Pfam data available for DKFZphfd2_46d13.1)

DKFZphfkd2_46j20

group: metabolism

DKFZphfkd2_346j20 encodes a novel 224 amino acid protein similar to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein seems to be the human ortholog of 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase.

The new protein can find application in modulating the homoprotocatechuate degradative pathway and as a enzyme for biotechnologic production processes.

strong similarity to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

complete cDNA, complete cds, EST hits,
potential start at Bp 16 matches kozak consensus ANCatgG
strong similarity to proteins of worm plant archea and bacteria
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase is part of
the tyrosine metabolism (degradation of tyrosine late step) EC 5.3.1.-
complete cds according to similar C.elegans and A.thaliana protein

Sequenced by MediGenomix

Locus: unknown

Insert length: 1706 bp

Poly A stretch at pos. 1686, polyadenylation signal at pos. 1667

```
1  CACTTGATGG GAATCATGGC AGCATCCAGG CCATTGTCCC GCTTCTGGGA
51 GTGGGGAAAG AACATCGTCT GCGTGGGGAG GAACACGCG GACCACGTCA
101 GGGAGATGCG CAGCGCGGTG TTGAGCGAGC CCGTGTGTTT CCTGAAGCCG
151 TCCAGCGCCT ACGCGCCCGA GGGCTCGCCC ATCCTCATGC CCGCGTACAC
201 TCGCAACCTG CACCACGAGC TGGAGCTGGG CGTGGTGATG GGCAGCGCT
251 GCCCGCAGT CCCCAGGGCT GCGGCCATGG ACTACGTGGG CGGCTATGCC
301 CTGTGCTTGG ATATGACCGC CCGGACGTG CAGGACGAGT GCAAGAAGAA
351 GGGGCTGCCC TGGACTCTGG CGAAGAGCTT CACGCGTCC TCCCGGTCA
401 GCGCGTTCGT GCCCAAGGAG AAGATCCCTG ACCCTCACAA GCTGAAGCTC
451 TGGCTCAAGG TCAACGGCGA ACTCAGACAG GAGGGTGAGA CATCCTCCAT
501 GATTTTTTCC ATCCCCTACA TCATCAGCTA TGTTCCTAAG ATCATAACCT
551 TGGAGAAGAG AGATATTATC TTGACTGGGA CGCCAAAGGG AGTTGGACCG
601 GTTAAAGAAA ACGATGAGAT CGAGGCTGGC ATACACGGGC TGGTCAGTAT
651 GACATTTTAA GTGGAAGAGC CAGAATATTG AGTTATTCTT TAACAAGTTT
701 CGAGAGAGAA GGGAGCAAGA CAAGAGCAAG CAACGGCTAT TAAATGTCAC
751 AATCCTTTAA TTAGAAACCA TTTATTGGCC GGACGCGGTG GCTCACGCCCT
801 GTAATCGCAG CACTTTGGGA GGGCGAGGCG GGGCGCTCAC GACGTCAGGA
851 GATCCAGACC ATCTTGGCTA ACAGGGTGAA ACCCGTCTCT TACTAAAAAT
901 ACAAAAAATT AGCCGGGCGT GGTGGCGGGC GCCTGTAGTC CCAGCTACTC
951 TGGAGGCTGA GGCAGGAGAA TCAATTGAAC CCGGAGGCGG GAGCTTACAG
1001 TGAGCTGAGA TTGCGCCACT GTACTCCTGG GCAACAGCGA GACTCCGCTC
1051 CAAAAAATAA AAAAAAATAA AGAAACCATT TATTTTAAAA ATGATTAGAT
1101 TGCTATGCCT CAACTCATAG AAGATGAACC CTCGAAGAAA ACGTGAAGTA
1151 GAACGGGTGG GCCAGAAATG AAAACAGGCA AGTAAAGTAT TTCTTCGGAA
1201 AACATTTTAT CAAACCAAAT GTTAAAAAGA CTTTCCTTTT GTAAAACTGG
1251 ATTAGAGAAG ACTTTTCAGT GGGTTATCTC TAGGATGATC AGTAGTTCAG
1301 CACTTAAAAA CTGCAGAGAA AACTGAAAGT TATGTTCCAG ATAACTTTCC
1351 GTTGTTTACC AAATTTTCTT AGATTGGTTC ATCATCAGGA AGCATTTGTA
1401 AAAATAAAAA TCTCCACAAA TTACTGGCCC ATCTCGGACT TGCTGAATCA
1451 ATTTGATAGG ATTAATCTCC AGTGAAGCTG TGTTCACAGG GCATTCCAAG
1501 TGATTCTTAT CAGGAAATGT GAAAAACACT CCTGTACATA ATCGGTTAAT
1551 TTAATAATTT ACTTAATAAG TGAACAAGTA ATGAAGATTT CACCTGTTTT
1601 CTTAGGGTAT CTACCCAGAC CCATCGATTG TGAGTTCGGG AGATGATTTT
1651 GAAATTACTG TTTTCCAAAT AAAGGTGCTC CTTTCCAAAA AAAAAAATAA
1701 AAAAAA
```

BLAST Results

No BLAST result

Medline entries

94039092: Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of *Escherichia coli* C.

Peptide information for frame 1

ORF from 7 bp to 678 bp; peptide length: 224
Category: strong similarity to known protein

```

1 MGIMAASRPL SRFWEWGKNI VCVGRNYADH VREMRSVAVLS EPVLFKLPST
51 AYAPEGSPIL MPAYTRNLHH ELELGVVMGK RCRAVPEAAA MDYVGGYALC
101 LDMTARDVQD ECKKKGLPWT LAKSFTASCP VSAFVPKEKI PDPHKLKLWL
151 KVNGLRQEG ETSSMIFSIP YIISYVSKII TLEEGDIILT GTPKGVGPVK
201 ENDEIEAGIH GLVSMTEKVE KPEY

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphfkd2_46j20, frame 1

PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*, N = 1, Score = 537, P = 8.7e-52

PIR:D71109 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase - *Pyrococcus horikoshii*, N = 1, Score = 529, P = 6.1e-51

PIR:C71425 hypothetical protein - *Arabidopsis thaliana*, N = 1, Score = 519, P = 7e-50

PIR:A64864 probable 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase b1180 - *Escherichia coli*, N = 1, Score = 474, P = 4.1e-45

>PIR:S44919 ZK688.3 protein - *Caenorhabditis elegans*
Length = 214

HSPs:

Score = 537 (80.6 bits), Expect = 8.7e-52, P = 8.7e-52
Identities = 99/211 (46%), Positives = 138/211 (65%)

```

Query:   10 LSRFEWGWGNIVCVGRNYADHVREMRSVAVLSEPVLFKLPSTAYAPEGSPILMPAYTRNLH 69
          L+ F      IVCVGRNY DH  E+ +A+ +P+LF+K  ++ EG PI+ P  +NLH
Sbjct:   4  LAGFRNLATKIVCVGRNYKDHALELGNALPKKPLMFVKTVNSFIVEGEPIVAPPGCCNLH 63

Query:   70 HELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWTAKSFTASC 129
          E+ELGVV+ K+  + ++ AMDY+GGY + LDMTARD QDE KK G PW LAKSF  SC
Sbjct:   64 QEVELGVVISKKASRISKSDAMDYIGGYTVALDNTARDFQDEAKKAGAPWFLAKSFDGSC 123

Query:   130 PVSAFVPKEKIPDPHKLKLWLKVNGLRQEGETSSMIFSIPYIISYVSKIIITLEEGDIIL 189
          P+ F+P  IP+PH ++L+ K+NG+ +Q  T MIF IP ++ Y ++ TLE GD++L
Sbjct:   124 PIGGFLPVSDIPNPHDVELFCKINGKQQRCRTDVMIFDIPTLLEYTTQFFTELVGDVVVL 183

Query:   190 TGTPKGVGPVKENDEIEAGIHGLVSMTEKVE 220
          TGTP GV  +  D IE G+  ++ F V+
Sbjct:   184 TGTPAGVTKINSGDVIEFGLTDKLNKSFNVQ 214

```

Pedant information for DKFZphfkd2_46j20, frame 1

Report for DKFZphfkd2_46j20.1

```

[LENGTH]      224
[MW]           24843.07
[pI]           6.96
[HOMOL]        PIR:S44919 ZK688.3 protein - Caenorhabditis elegans 8e-55
[FUNCAT]       r general function prediction [M. jannaschii, MJ1656] 9e-40
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YNL168c] 4e-38
[EC]           5.3.3.10 5-Carboxymethyl-2-hydroxymuconate delta-isomerase 1e-35
[PIRKW]        isomerase 1e-35
[PIRKW]        intramolecular oxidoreductase 1e-35
[SUPFAM]       2-hydroxyhepta-2,4-diene-1,7-dioate isomerase 1e-46
[PROSITE]      MYRISTYL 4
[PROSITE]      AMIDATION 1

```

[PROSITE] CK2_PHOSPHO_SITE 2
 {PROSITE} PKC_PHOSPHO_SITE 3
 {KW} Alpha_Beta

SEQ MGIMAASRPLSRFEWGNIVCVGRNYADHVREMRSAVLSEPVFLKPSTAYAPEGSPIL
 PRD cccccccccchhhhhccceeeecchhhhhhhhhccccceeeeccccccccccccc

SEQ MPAYTRNLHHELELGVVMGKRCRAVPEAAAMDYVGGYALCLDMTARDVQDECKKKGLPWT
 PRD cccccchhhhhheecccccccchhhhhhhheeeecchhhhhhhhhhhcccccc

SEQ LAKSFTASCPVSAFVPKEKIPDPHKLKLWLKVNGLRQGETSSMIFSIPYIISYVSKII
 PRD cccccccccceeeecccccccccceeeecccccccccceeeecchhhhhhhhh

SEQ TLEEGDIILTGTGPGVGPVKENDEIEAGIHGLVSMTFKVEKPEY
 PRD hccccceeeecccccccccceeeeccccccccccccc

Prosite for DKFZphkd2_46j20.1

PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	192->195	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00008	2->8	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	116->122	MYRISTYL	PDOC00008
PS00008	191->197	MYRISTYL	PDOC00008
PS00009	78->82	AMIDATION	PDOC00009

(No Pfam data available for DKFZphkd2_46j20.1)

DKFZphfkd2_46k19

group: transcription factors

DKFZphfkd2_46k19.3 encodes a novel 130 amino acid protein similar to rat Dcoh, a bifunctional protein-binding transcriptional co-activator.

Dcoh is a bifunctional protein, complexed with bipterin. It serves as dimerization cofactor of hepatocyte nuclear factor-1 and catalyzes the dehydration of the bipterin cofactor of phenylalanine hydroxylase.

The new protein can find application in modulating/blocking the expression of genes controlled by the hepatocyte nuclear factor-1.

strong similarity to pterin-4-alpha-carbinolamine dehydratase

potential start at Bp 102 according to similar proteins,
both genomic sequences are from chromosome 5,

Sequenced by MediGenomix

Locus: map="5"

Insert length: 5641 bp

Poly A stretch at pos. 5617, polyadenylation signal at pos. 5598

```
1 CAGCCCTCGG CAGACGGCCA ATGGCGGCGG TGCTCGGGGC GCTCGGGGCG
51 ACGCGGCGCT TGTTGGCGGC GTCGCGAGGC CAGAGCCTAG GGCTAGCGGC
101 CATGTCTACA GGTACTCACA GTTGATTGCG AGAGGAGAGG AACCAAGCTA
151 TACTTGACCT TAAAGCAGCA GGATGGTCGG AATTAAGTGA GAGAGATGCC
201 ATCTACAAAG AATTCTCCTT CCACAATTTT AATCAGGCAT TTGGCTTTAT
251 GTCCCGAGTT GCCCTACAAG CAGAGAAGAT GAATCATCAC CCAGAATGGT
301 TCAATGTATA CAACAAGGTC CAGATAAATC TCACCTCACA TGACTGTGGT
351 GAACCTGACCA AAAAAGATGT GAAGCTGGCC AAGTTTATTG AAAAAGCAGC
401 TGCTTCTGTG TGATTCTCTC CAAAATACAT AAGTCTGAGA GGCTAACTTT
451 GATGGCTGTG TTAACATATG TCACGTGTAG CACAGTGGAG AAAGCAGGAT
501 ATGGCTCATA ATGACAGTGG TGAAGACCTG CGAATGAAGT TGCTAGTTAA
551 CACCTACATT AGGGTTTGAC ATAGGTCTAT GTTATGGGTC GCTGCATCTG
601 CTGGAACCTA CAGACTTTAC TATAGAGAAAT CAAAGATCCC GTATCCGAAG
651 TCTATGGAAA TGCTCATGGT GGTAAATTC AACAGAATGA AACACCAAAAC
701 TTGCTTAAAG TAACACAGT TTCAATTTGA AAGAGATATT GTCAAAATTG
751 GAGGCCCCCA GGTTCCTGTC TGTTCCAAAT CTTTGCAATG TGACAGTGGT
801 TTCTCTGATG TGGTAAGCTT TGGCTTTCTT CTGTTTTCTT TCTAAAAGAT
851 CACTGGAGTA GAGAGGAGTT AAACAGACAT GACCTTTGAC CTCTTGATG
901 ACCTCCACAG ATAGCAAACC GGGCCGACAC ATGGTTGACG ATGCTCTTTT
951 CTACAATGAA GTTAATGAAA GTTCTGAAAA TAGTGATTAC TTTCTGACAT
1001 TGATAGGATT TAGGAAACCT CTGGATAAAT AGCTTAAGCA TGGCTGTTTA
1051 TGTTTTTGCT ATAGACAAA AGCAGCAGCA TGTACATTGT ATTTGGACAC
1101 AAGCCTGCCT CGGTTAATAT ATTGAACAT TGGACCACTA GGGTTAGTAG
1151 GGAGCGGCT GTACACTTTT TGATTACGCA TTCAGAAACA TTCTAGGTGG
1201 ACTCTGTAGC TTTCAAGTTT GTAAAGTTAT CGGAAAACA TCGGAGGGGT
1251 TTGGCCATCA TATGTAGCT TTCTGTTTCA ATGCCACTTA CTCAGGATTA
1301 GTAATTAAT GACTGTCCAG AGGACTTCAG GGTACCAAG CTGCTGCACC
1351 TGCCATTGGC TGACTCTCCC CGGCTATCTG TGGCTGAGAT GGTGCTGCTT
1401 AGGTCACGCA GAGCATGAGC TGCTGCTGAA AGGGCACAGG AGATGGCCCT
1451 TGGGCTTCTC ATCCCAGGAT GCCTGCCCTG CCCACCAATC CATGAGAAGA
1501 TATGTATGAT TTCAGTAGGC CTTGGATCAG CTTGTCACCT CTGGTTTCTT
1551 GTTTGCTTTT CACTCACTCA GCTGGAGTTT CATTCCAGA CTAAAGTCTT
1601 CATCATTTGC TTCAGAAACA GCATTATCTT GTGGCTGTGC TGATGTAGTA
1651 CACCAAGAAC AACTGGGCTC TTCTCTGTCA CTTTCAGTGG GCTACCTTCC
1701 CTCACCTCTC CAAGCAGCAT GAAAGAATTC TTTACATTTT TAATCTCTTT
1751 TTTGTTTTTC CCTGAAAGTA TGCTTTGGTG CTTAAAGAGA GAAGTCACAA
1801 AAGTATACTA CTGAGTTTCC TGGAGTGA ATCCTGTTGT CCCTAGCTAT
1851 GTGAATGAGC ACAGGGATCC CTGATGCCAT TATTTTGTAT ATTCATACGG
1901 CACACACTTA CTGAGGGCCT TCTGTGTGCC CTAGGGGATT GAGCACAGTG
1951 ACATATCAGG GCAGGTAGAA ACAGATGGAG AGCTGATGCG GGCTGTCTTA
2001 GAGCAGCTGC CCCAGGAGGC CCCTGTGGAT GGATGTTGCG CAGGAGCCCT
2051 GAGACGTTAG GGGCATATAA CTAAGGACA TAGCAGGAGT TATAGGAGGA
2101 GCTGATCCCT GAGGGAACA ATGAAGACGG AGAAGATGGG GCTAAAGTTT
2151 GAATTGTGGG GACATTAATC ACGGTGATTC TTAAGACTTT GCTGTTGATG
2201 ATTTTAAATG GAGAAATGA GTACGTAGA TGTATTTC CAGTTCAGTA
2251 TATAGGTTGC CCACAAAGTA TTTTCTACC ATGAATGGTC ATATATACTT
2301 GTTGTAGAAT ACCAGGGACA GCAGAGATGG TGGGGTAGTT ACTTCTTTT
2351 CTTACAGCCC AAGAACTTTG GTGTCCAGGA GATTGACCAA TTTAGCCACT
2401 GAGCATTTAA TACAACACAG GGCTACCCAG ATCCCACTGT CTTGATTGCG
2451 CCGTAAAGCC AAGGGAGTCA GGAGAAGGTG AGTGGGGTGA ATATATTAAT
2501 CCTGAGAGTT GAACAGAGCA AAAATCCCTA TTACTTTTGT ACTTAAAAACA
```

```

2551 TCTCTGCCAC ATGTGCTCAC TCTTTATATT CTGTTTAGGT GGTATATATG
2601 TGCACATCCC ATCCTATGCC TGCAGTTAGC CAACTCAGGG TTTATATTGC
2651 CTCCTTTCTT TTTTCTTTT TTTTTTTTTT TTTTAAGAGA TGGGGTCTCG
2701 TTCTGTCTAG CAGACTGGAG TGCAGTGGTG TGATCACAGC TCATTGTAAC
2751 CTCACACGCC TGGACTGAAG TGATCCTCCT GCCTTGGCCT CTCTGGTAGC
2801 TGGGACTACA GGTGCATGCC ACCACACCCA CTAATTTTTT TTTATTTTAA
2851 TTTTGTGTAG AGACAGTCTC ACTATCTTGC TGGGGTGGT CTTGAACCTC
2901 TGGGCTCAAG TTATCTTGCT GCCTCAGCCT CCCATGGGTA ATCTTTATTT
2951 CCTTTTTTTT TTTTTTTTGG AGATGGAGTT TCGCTCTTGT CGCCAGGCT
3001 GGAGTGCAAT GGCACGATCT TGGCTCACTG CAGTCTCCAC CTCTGGGTT
3051 CAGGTGATTG TCCATCTCG GCCTACTGAG TAGCTGAGAT TACAGGCAAC
3101 TGCCACCATG CGCGGCTAAT TTGTGTATTT TTTTGTAGTA AGAGATGGGG
3151 TTTCGCCATG TTGGCCGGAC TGGTCTTAGA CTCCTGACCT CAAGCGACCT
3201 GCCTGCCTTG GCCTCCCAAA GTGCTGGGAT TACAGGCATG AGCCGCTATG
3251 CCTCGTCGCT GATTTTATT TCTTATTTT TTTTAGAGA TGGGGTCTC
3301 ACTATGCTGC TCAGGCTGAT CTCAAACCTC TGGCCTCAAG TGATCCTCC
3351 ACCCTAGCCT CCAAGTTGC TGGGATTATA AGTGTGAGCC ACTATCCCTA
3401 CCTCACTATT ACCTTCTTTG CTCTCTTGT TTTCTTTGT TCTAAGTCAA
3451 ACCCATCACA ATCTTTCTT GTCTTCCAG GTGTTTCCA GTGCTGTGCC
3501 CTGGATGTGC TCTCTTCTC TTAGAGCCCA GAGAACTGCT TTTTCCCTC
3551 TATATATGAC CCTTAACCTT TTCTAACACA TTATTAAGGG CCTGTGCTA
3601 TCAGCTGGGG GCACCTCTTG AAGGGAGGGC CTTTGTGTGG TCTGTTCTA
3651 GTGACTTCCA GCTTTAACC AGAGCCTCAT GATTGCTGGG TGCCATAGC
3701 CTTTGTGCTG AATGGAGGCA CTCACCTCC TTTGGAAGAG AGAATCCATG
3751 ATAGACCCAC TTGGGAGCTC CCCACTTCAG GGGCCTACAC ACTGTAATG
3801 CAACAGAATG CCCAAGAGTG ACCTCATAAA GCAAGGATTG CCTTCGTGGC
3851 CCCTTCTCTG CTGCTCTCA GAATCCAGAC GCTAAGGAAA ATCCCTAAGC
3901 AGAGATTTC TGTGGATGC TAAAGCAAG GAATAAAGT TGAATTTTG
3951 GAAATGTCT CAACACGCTC ACCAGCGCCA CTCGAGAGTC ATTTCTAGTT
4001 CACCAAGTTG CACTACATCG GTGGGATTTT GCCCAACATT CAAGAAATTT
4051 AAGTAAATAT TATCTATCTC CATTGCCTGT TAAGAAATGT GCTAGTAGAA
4101 GTGTGAGGGC AGGGTGTGAG TGTCTCTCA GCCTCTTCCC TCAGATACTC
4151 GTCTGTCTAC CAAATAAGT TGATGTCTT TGACAACTGT GTTCTATGA
4201 TTGGTGAGGC TGGCATGCTA TTACCTTTAT GTGCCCTGTA GACTTGAATG
4251 ACCAGTTTGA CCAAGTTGAC TGTAGATAA TCAGAAAGCT TTTCTCTTTT
4301 TTTATAATAG ACCCATCTC AAATCAGATA ATGAAATTA CATATCTTGA
4351 TATATTAGAA AAGTATATAC ATTCTGGCTG GGCACGGTGG CTCACGCCCTG
4401 TAATCCCTGC ACTTTGAGAG GCTGGGGCGG ATCACTTGAG GTCAGGAGTT
4451 TGAGACCGGC CTGGCCAGCG TGGCGAAACC CCATCTCTAC TAAATAACA
4501 CAGATTAGCC CGGAGTGATG GTGTGCACCT GTTGCTCCAG CTACTCAGGA
4551 TGCTGAGGCA GGAGAATCCC TTTAACCTGG GGGCGAAGG TTGCACTGAG
4601 CCAGGATTGC ACCACTGCAC TCCAGCTGG GTACCGAAG GGCCTCTGT
4651 CTCAGAAAAA AAAAAAAGA AGAGGAAAAA GAAAAATATA TATTCTATAT
4701 TTTTAACTAT TATGAGAATG TGTTCATTTC ATTTGTAACA TAAATGGGAA
4751 AACCAATAA CTAATCTGTA GAAAAATTC AAAGCACAGA TAAATGGGAA
4801 TAAACAGGAA AAAGAATCAC CTATAACCTC ACCATCCATA GACAGACACT
4851 GTTAAATTTT TGGCATAATT CCGTCTGATT TTTTCTACTG CTGATTTTGT
4901 CACAGGTGAG ATAATTTTGA ACAGAGAATT TTGTATCTTT GGTTTTTTGTG
4951 TTTCTGCTGA CACAAAAACA AAAGATATAA AAATGGATCA TAAACATTTT
5001 TCTAAATCCT GAAAGTGCA TAGACATATT TTAGTGCTGT TATTTCACAA
5051 GATGGACATA CCATAATTTA CTTACACAGT CCTTTTGTG AGATGTTTAA
5101 GTTGTTTTCA AGCTTCTCAG TGCTGGAAAA AATACTGAGA TAGACATGTT
5151 TAGTTGAAGT TATTTTATTT CAGGTTATAT TATCTTGGGT CAGAGAATGA
5201 ATGGTTCTCA GGCTTTTCAA AAGAGCTGGT CAGTTTTAT GCCTCTGGCA
5251 GTTTTGTAGA GTGCTCAATC ATACTACACT GTTGCCAGCA TTAGATCTTA
5301 TCACATTAA GTCAATGCTA ATTTTATAAA CAAAAACAAT GGTTTTACTT
5351 TGCACTCTCC TGATTGGTGT TGCTGTAGAA CATATTGGA GAAGTTTGT
5401 TGCTTTTGGT GTTTATTCCA TGAATAGATT GTGTGCCAT TTTCTTTGG
5451 GGTATTCACT TTTTATTAC TGATGTGAGC ATGTGTATGG GTGATTATTT
5501 GATGATTATC AGTTTGTGCT AGTAGACTGG CAATATTTAG TCTTGTCTG
5551 ACTGTGTTCC CAGTGCCAAC TAGATTGCTT GATATGTAGT TGCCACTCAA
5601 TAAAGATTGG TTGAGTCAAT GAAAAAATAA AAAAAAATAA A

```

BLAST Results

Entry AC004764 from database EMBL:
Homo sapiens chromosome 5, P1 clone 255g5 (LBNL H61), complete
sequence.

Score = 11057, P = 0.0e+00, identities = 2217/2224
Bp 428-5625 of cDNA == Bp 2912-8107 of AC004764

Entry HSAC1555 from database EMBL:
Homo sapiens (subclone 1_d8 from BAC H75) DNA sequence, complete
sequence.

Score = 575, P = 5.1e-30, identities = 115/115
Bp -240- 430 of cDNA == HSAC1555 splice pattern

.....

95242099:
Crystal structure of DCOH, a bifunctional, protein-binding
transcriptional coactivator

1 MAAVLGALGA TRRLAALRG QSLGLAAMSS GTHRLIAEER NQAILDLKAA
51 GWSELSEDA IYKEFSFHNF NQAFGMSRV ALQAEKMNHH PEWFNVYNKV
101 QITLTSHDCG ELTKKDVKLA KFIKAAASV

Report for DKF2phfkd2_46k19.3

```

SEQ      MAAVLGALGATRRLLAALRGQSLGLAAMSSGTHRLTAEERNAOAILDLKAAGWSLSERDA
SEQ      .xxxxxxxxxxxxxxxxxxxxx.
1dchb    .....CCSCNNNNNNNNNNNNNNNNHCCCEEECCCE

SEQ      IYKEFSFHNFNQAQFGMSRVALQAEKNHHHPWFNVYNKVQITLTSHDCGELTKDKVKLA
SEQ      1dchb    EEEEECCCHNNNNNNNNNNNNNNNNNNHCCCEEEETTTEEEECBTTTBTCCNNNNNNNN
SEQ      KFIKAAASV

```

SEG
1dchB HHHHHHHHHH

Prosites for DKFZphfkd2_46k19.3

PS00005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS00005	32->35	PKC_PHOSPHO_SITE	PDOC00005
PS00005	56->59	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113->117	CK2_PHOSPHO_SITE	PDOC00006
PS00008	6->12	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2_46k19.3)

DKFZphfkd2_46m4

group: signal transduction

DKFZphfkd2_46m4.3 encodes a novel 198 amino acid putative GTP-binding protein related to the SAR-1 family of Ras superfamily members.

SAR1 proteins are involved in vesicular transport between the endoplasmic reticulum and the Golgi apparatus.

The new protein can find clinical application in modulating the transport of vesicles to the Golgi Apparatus, thus enabling post-translational modifications of the vesicles contents. Blocking of the molecule is expected to result modulation/blocking of secretory pathways.

nearly identical to mouse GTP-binding protein

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: /map="438.9 cR from top of Chr10 linkage group"

Insert length: 2996 bp

Poly A stretch at pos. 2969, polyadenylation signal at pos. 2958

```
1 ACATCCGGCG AGTAGCTGGC GGTCCCGGGT GCTGCTGGTT AGTGTGCTCT
51 GAGGGAGGGT CCGAGCCAGC CGCTGTTTTG CCGGAGGAGC CCCTCAGGCC
101 GTAGTAAGCA TTAATAATGT CTTTCATCTT TGAGTGGATC TACAATGGCT
151 TCAGCAGTGT GCTCCAGTTC CTAGGACTGT ACAAGAAATC TGGAAAACCT
201 GTATTCTTAG GTTTGGATAA TGCAGGCAAA ACCACTCTTC TTCACATGCT
251 CAAAGATGAC AGATTGGGCC AACATGTTCC AACACTACAT CCGACATCAG
301 AAGAGCTAAC AATTGCTGGA ATGACCTTTA CAACTTTGA TCTTGGTGGG
351 CACGAGCAAG CACGTCCGGT TTGAAAAAAT TATCTCCAG CAATTAATGG
401 GATTGTCTTT CTGGTGGACT GTGCAGATCA TTCTCGCCTC GTGGAATCCA
451 AACTTGAGCT TAATGCTTTA ATGACTGATG AAACAATATC CAATGTGCCA
501 ATCCTTATCT TGGGTAACAA AATTGACAGA ACAGATGCAA TCAGTGAAGA
551 AAAAATCCGT GAGATATTG GGCCTTATGG ACAGACCACA GGAAGGGGGA
601 ATGTGACCCCT GAAGGAGCTG AATGCTCGCC CCATGGAAAT GTTCATGTGC
651 AGTGTGCTCA AGAGGCAAGG TTACGGCGAG GGTTCGCGCT GGCTCTCCCA
701 GTATATTGAC TGATGTTTGG ACGGTGAAAA TAAAGAGCTT TACTTCTCT
751 GGACTGATCC TATTCACAGC TTCCTCATGA ACTTTTCTAA TAGAACAAAG
801 ATAGCTCTCC AACCATGTCT GGCCTTGAGA AGCCAAGAGT CTCTGTCAAC
851 TCTCTCATTG CCCAGTGGTG ACATGTGCTC TTCTCCACAC TGTGGGAGG
901 TAATGCTGCC CCACGTGCTG GTGCAGGTCA GTATCCTGGG ACTTGGAAAG
951 TGCGAGGATT TGCCGGGTAA AGCTGTATGC CATCATGGGG CACCTGAAAA
1001 GAAAAACACG TCTCACCACG GTGGTTGATT CAAAAGAAAG TGATTCTATT
1051 TTTTAAAGAA AGCGTTGTTA ATGTAATTGG TATCCCTCCT AACTTTTTGA
1101 GTTCACAATT TACTTGGTCC AGAGTTTTCT ATTCTTTTTT TTTTTTAA
1151 CTAATGAATG ACATTAGAT ACTTCATAAA ATTATGAACA GATATGGAGG
1201 CCAGAGCTCA TTTGGGTAAA CTTACTCCTG CTGAGTTAGC AGGTTGGTGA
1251 GAGAAGCTCC CCGAGCTCA CCGTCTCTC TGACTGCCTT GGAGTAGGTG
1301 GCATAACCTT GTGCACAGAG AACTAGAAAA GGGGAGAAAC CCGGCGCTTG
1351 CAGTTGTGGC AGGTTTCCAC TGTGGTAAGC TAGGTTTATT CCTCATCAAG
1401 GAATGTGTAG CAGATTGTTT ACTGTGGAGG AGGTAATTAT AGAATGGGTT
1451 ATTGTTGTTA TTCTTACTCA TGAAGTTACA GATTTTAGCC AGTCTTTGCT
1501 TTTATATCTT TGTGAAATTT AATTCTCTCT TATAGCACCT TCCTTTTTCG
1551 TTTTCAGTTA TCAAAAGTGA CTTTGACCTC ATAAGAGAGT TGAGAACATC
1601 TCTCGTGCTA CATACTGCAG GTGCATCAGT TACTTTTCCA CAGATTCTAG
1651 GGGGACATTT TTCTGAATAG GAAGACAGGA CAAAGTTAAG AGCTTAAGGG
1701 CTCTTAATTC TGTGAGTTGA GGACTTAAAA GTATTGTAGC ATTTGTTTGG
1751 ATCCATGAAA AATGTATTCA GTGGGCTTTA AAATTTCCAT TTGCAGAAAT
1801 TGGTCTCTCA GGCTGTTTGG GAGCTCTTTT TTTTACATT TTTCTCCTTT
1851 GACACCTATT TTATTGGTGT TTAAGTAAA GGTAAACATC TGTAGCTTTT
1901 CCAGGTTTTT TTTTTTTTTT TTGATATGAA ATTGTCTTTC TCCATTGCAG
1951 AAATAAGCTA GGGAAACACT AACCCAAAAA CTTTCTGTAG AGCTGTTCCCT
2001 TTGGAGGCAG CATCACTTAT TGGCAGTAAA GACTCAGTAT AAAAGCACCA
2051 GCATCCCTAC TTGGGTGATG GGGATTAATT TTATAGCATT CCATTTTCCT
2101 AGTGCCACAT GTGAAATTGG ATTTTGATGA TCTTAATCTA TATTCTACCC
2151 TTATAATAAA AGATCAAAAG ATATATCTCC TATGAACAGA TTGGAGATAG
2201 GAGATGAAAA GTTGGGAGGA TGCCTTTATT CTAATGTGAG GGTAGGGAAA
2251 ATGTGGATAA CATTACTGGG GTGAAGGAGG CATTGTCTCT TAGTGGAGT
2301 TCTCATTTTT ATTCTCCAGT ACTGACTTGT GGGGAAAGCA TACTTTTCA
2351 CTGCCAGGTA CTGAATGCAG AGGCTCAGTG AAGTATATAT GTGGGAAGTG
2401 CATGCATTTT GTTTATTAGC AAACATAGCT GGATTAAGAC GAAAGTTGTT
2451 GTTTGGAAAG GGGTTAAAGC CTTAAGTGAA CAATCTAGC TAACAGTGAA
2501 TGAAGTAGGT AATATAACTT GCATATTTTT AATTCCTTTT GGTTAAAGGT
2551 CCCCATACT TCTCTGTTCC GAGACATGAG AAGTATGATT ACTTCAGTGT
```

```
2601 TAGTTTCTT AATTTTTTTT TTCCCTATT TGTCCTTGT CACTTTGTTG
2651 CAAGCTAGAA ATCTGTGGT TATACATAGG GCAGCTCTTT GCGAAAGTGG
2701 TTTATTCAC TGGAGAAAGG GGATTGAAA TCAGTTAGAA CCAATGTATT
2751 TCTTGCCCA CGGACACTA TTCTATAAG ATAGCTGAAA GAAGCTGCTG
2801 TGAGGAGCTC AGCTCCAACA CAGGATCAGC ACCTGTGATA GGAATTCCCA
2851 TGAATTATGA CTCTCATTC TGTTTATCA GAGTGCATAT ATGCTCTACT
2901 TCAGGAAAAG TAAACAGTC ATTTACGAAA GAAAGTCAAT CTGTATCCTA
2951 AGCATTTTAA TAAAAGTTA AAACAAAAA AAAAAAAAA AAAAAA
```

BLAST Results

Entry HS679348 from database EMBL:
human STS WI-16722.
Length = 265
Minus Strand HSPs:
Score = 1242 (186.4 bits), Expect = 2.8e-50, P = 2.8e-50
Identities = 260/265 (98%)

Medline entries

94085558:
Molecular analysis of SAR1-related cDNAs from a mouse
pituitary cell line.

Peptide information for frame 3

ORF from 117 bp to 710 bp; peptide length: 198
Category: strong similarity to known protein

```
1 MSFIFEWIYN GFSSVLQFLG LYKKSGLVF LGLDNAGKTT LHMLKDDRL
51 GQHVPTLHPT SEELTIAGMT FTFDLGGHE QARRVWKNYL PAINGIVFLV
101 DCADHSRLVE SKVELNALMT DETISNPIL ILGNKIDRTD AISEEKLREI
151 FGLYGQTTGK GNVTLKELNA RPEVFMCSV LKRQYGE GF RWLSQYID
```

BLASTP hits

Entry S39543 from database PIR:
GTP-binding protein - mouse
Length = 198
Score = 1029 (362.2 bits), Expect = 5.1e-104, P = 5.1e-104
Identities = 197/198 (99%), Positives = 198/198 (100%)

Entry SARA MOUSE from database SWISSPROT:
GTP-BINDING PROTEIN SARA.
Length = 198
Score = 1012 (356.2 bits), Expect = 3.2e-102, P = 3.2e-102
Identities = 195/198 (98%), Positives = 196/198 (98%)

Entry CEZK180.4 from database TREMBL:
gene: "ZK180.4"; Caenorhabditis elegans cosmid ZK180.
Length = 193
Score = 679 (239.0 bits), Expect = 6.3e-67, P = 6.3e-67
Identities = 125/197 (63%), Positives = 161/197 (81%)

Alert BLASTP hits for DKF2phkd2_46m4, frame 3

No Alert BLASTP hits found

Pendant information for DKF2phkd2_46m4, frame 3

Report for DKF2phkd2_46m4.3

{LENGTH}	198
{MW}	22367.00
{pI}	6.21
{HOMOL}	PIR:S39543 GTP-binding protein - mouse 1e-112


```

[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YPL218w]
1e-58
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YPL218w] 1e-58
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR094w] 2e-23
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YPL051w] 4e-22
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YDL192w] 3e-20
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 3e-19
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 2e-09
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 2e-09
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YHR168w] 7e-05
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YKL154w]
1e-04
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YHR005c] 1e-04
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YKL154w]
1e-04
[FUNCAT] 08.19 cellular import [S. cerevisiae, YML001w] 3e-04
[BLOCKS] BL00395A Alanine racemase pyridoxal-phosphate attachment site proteins
[BLOCKS] BL01019B ADP-ribosylation factors family proteins
[BLOCKS] BL01019A ADP-ribosylation factors family proteins
[BLOCKS] BL01020D SAR1 family proteins
[BLOCKS] BL01020C SAR1 family proteins
[BLOCKS] BL01020B SAR1 family proteins
[BLOCKS] BL01020A SAR1 family proteins
[SCOP] d1plj_ 3.25.1.3.1 CH-p21 Ras protein [human (Homo sapiens)] 7e-36
[SCOP] d1guaa_ 3.25.1.3.10 Rap1A [Human (Homo sapiens)] 8e-40
[SCOP] d1rrf_ 3.25.1.3.5 ADP-ribosylation factor 1 (ARF1) [rat (Rattus)] 2e-55
[SCOP] d1hurb_ 3.25.1.3.4 ADP-ribosylation factor 1 (ARF1) [human (Homo)] 1e-58
[SCOP] d1gota2_ 3.25.1.3.3 (1-54,171-326) Transducin (alpha subunit) [ra] 2e-33
[SCOP] d1tadb2_ 3.25.1.3.2 (1-30,152-316) Transducin (alpha subunit) 6e-36
[PIRKE] glycoprotein 4e-19
[PIRKE] monomer 1e-16
[PIRKE] P-loop 3e-64
[PIRKE] lipoprotein 4e-19
[PIRKE] GTP binding 3e-64
[PFAM] ADP-ribosylation factor 5e-22
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 3
[PROSITE] SAR1 1
[PROSITE] CK2_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 3
[PROSITE] ASN_GLYCOSYLATION 1
[PFAM] ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW] Alpha_Beta
[KW] 3D

```

```

SEQ MSFIFEWIYNGFSSVLQFLGLYKSGRLVFLGLDNAGKTTLLHMLKDDRLGQHVPTLHPT
lhurA .....TTTTCCCCCEEEETTTTCHHHHHHHHCCCCCEEEETTEE

```

```

SEQ SEELTIAGMTFTTFFDLGGHEQARRVWKNYLPAINIGIVFLVDCADHSRLVESKVELNALMT
lhurA EEEEEETEEEEETTTTTCCHHHHHHCEEEEEETTTTTHHHHHHHHHHHHH

```

```

SEQ DETISNPVILILGNKIDRTDAISEEKLREIFGLYGQTTGKGNVTLKELNARPMVEVFMCSV
lhurA TTTTTTEEEEEETTTTTCCHHHHHHHC.....

```

```

SEQ LKRQGYGEGFRWLSQYID
lhurA .....

```

Prosite for DKF2phfd2_46m4.3

PS00001	162->166	ASN_GLYCOSYLATION	PDOC00001
PS00005	25->28	PKC_PHOSPHO_SITE	PDOC00005
PS00005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00006	60->64	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	111->115	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00017	32->40	ATP_GTP_A	PDOC00017
PS01020	171->197	SAR1	PDOC00782

Pfam for DKF2phfkd2_46m4.3

HMM_NAME	ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)		
HMM	*GMgWfsIFrkmWGIWNKEMRILMLGLDNAGKTTILYMLKlgEIVTTIPT		
	++	FS+++++GL++K+++++LGLDNAGKTT+L+MLK++++	+++PT
Query	9	-YNGFSSVLQFLGLYKKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPT	56
HMM	IGFNVETVeYKNIKFNVDVGGQdsIRPYWRHYpNTDGIWVVSaDRD		
	+++++	E++++	+++F+++D+GG++++R++W++Y P+++GI+++VD+AD++
Query	57	LHPTSEELTIAGMTFTTDFLGGHEQARRVWKNYLPAINGIVFLVDCADHS	106
HMM	RMeEaKqELHaMLNEEELrDAPILIFANKQDLpgAMSeSIREaLGLHeI		
	R+	E+K+EL+A++++E	++++P+LI++NK+D+ +A+SE++RE+ GL+ +
Query	107	RLVESKVELNALMTDETISNVPILILGNKIDRTDAISEEKLREIFGLYGG	156
HMM	RCn.....RPWYIQMCCAVtGEGLYEGMDWLSNYInkrK*		
	+++	RP++++MC+++++G++EG++WLS+YI	
Query	157	TTGKGNVTLKELNARPMEVFMCsvLKRQGYGEGFRWLSQYI-----	197

DKF2phfkd2_47a4

group: transcription factor

DKF2phfkd2_47a4.1 encodes a novel 280 amino acid protein with similarity to zinc finger proteins.

The new protein is a putative transcription factor with one C2H2 zinc fingers.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to C.elegans F46B6.7

potential frame shift at 1092, will be checked see BLASTX

Sequenced by MediGenomix

Locus: map="7q31"

Insert length: 1756 bp

Poly A stretch at pos. 1737, no polyadenylation signal found

```
1 CCCTTTTCTT TTCTGCCGGG TAATGGCTGC TTCCAAGACC CAGGGGGCTG
51 TCGCCCGAAT GCAGGAAGAC CGTGATGGGA GCTGCAGCAC AGTCGGGGGT
101 GTAGGTTATG GGGTAAGGAT TGTATCCTGG AGCCGCTTTC CCTGCCAGAA
151 AGTCCAGGTG GCACCACCAC TTTAGAAGGT TCTCCATCTG TGCCTTGAT
201 TTCTGTGAA GAACATTTTC CTGTGGCTGA ACAAGACAAA CTTCTGAAGC
251 ACATGATTAT TGAGCATAAG ATTTGTCATG CTGATGTCAA GTTGGTTGCT
301 GATTTCCTAAA GGTACATTTT ATATTGGAGG AAAAGGTTCA CTGAACAGCC
351 CATCACAGAT TTTTGTAGTG TAATAAGAAAT TAATTCACCT GCTCCATTTG
401 AAGAACAAGA GAATTATTTT TTGTTATGTG ACGTTTACCT AGAAGATAGA
451 ATTCTTAGAG AAGAGCTTCA GAAACAGAGA CTGAGAGAAA TTCTGGAACA
501 ACAGCAGCAA GAACCAAAATG ATAACAATTT TCATGGCGTT TGTATGTTTT
551 GCAATGAAGA ATTCCTTGGA AACAGATCTG TTATTTTGAA CCACATGGCC
601 AGAGAACATG CTTTCAACAT TGGATTGGCA GACAACATTG TAAACTGCAG
651 TGAATTTTTG TGTACATTAC AGAAAAAGCT TGACAATTTG CAGTCTTGT
701 ACTGTGAGAA GACCTTCAGG GGCAGAAATA CACTTAAAGA TCACATGAGG
751 AAAAAACAGC ATCGTAAGAT TAATCCTAAG AACAGAGAAAT ATGACAGATT
801 TTATGTCATC AATTATTTGG AACTTGGAAA ATCGTGGGAG GAAGTTCAGT
851 TGGGAAGATGA TCGGGAGTTG CTGGACCATC AGGAAGATGA CTGGCTGAT
901 TGGGAAGAAC ACCCTGCCTC TGCAGTCTGC TTATTTTGTG AAAAGCAAGC
951 AGAAACAATT GAGAAGTTGT ATGTCCACAT GGAGGATGCA CACGAATTTG
1001 ATCTTCTCAA AATAAAGTCA GAACCTGGAT TAAATTTCTA TCAGCAAGTG
1051 AAACCTGGTCA ATTTTATTCG GAGGCAAGTT CACCAATGCA GATGATGGCT
1101 GCCATGTGAA GTTCAAATCC AAAGCAGACT TAAGAACTCA CATGGAAGAA
1151 ACTAAACACA CTTGCTGCTC CCCGATAGA AAGACGTGGG ATCAACTGGA
1201 GTATTATTTT CCAACCTATG AAAATGACAC TCTCCTGTGT ACACATCTG
1251 ACAGTGAAAG TGACCTGACA GCTCAGGAAC AAAATGAAAA TGTTCCTATC
1301 ATCAGTGAAG ATACATCTAA ACTGTATGCT TTGAAACAAA GCAGTATTTT
1351 GAACCAAGTTG CTACTATAAG AGTACTTGAA AACCTAGAAG AAACCTACCAC
1401 AGAAGCAATT TTTTATGTTT TTCTCCTATG AGACAGATAT GAAAGAACAA
1451 TTTAAATTTG AACATCAACA AAAGATTGGT CTTGGTGAA ATAACTTTT
1501 CAAAAATGAA TGTCTTTTTC AAAAAATAAA GTAGAAAAAT GCACCTACTA
1551 AGAACATGAA AAAAAATGA AGTAGGAAAA TAAGATGAAG ACTTTGTATT
1601 TTGGCTGTAA AGTTTATTG TGTGATCATC TTAATATTAT TCACCTCATT
1651 AAACCTATAA TTATATATAG AAGTATATGT CAATTACAAA GAAATGAAAT
1701 GTTCAAAATTA TTTATAAACC TGATTTTTC AATCAGCGAAA AAAAAA
1751 AAAAAA
```

BLAST Results

Entry AC004112 from database EMBL:
Homo sapiens BAC clone RG313E03 from 7q31, complete sequence.
Score = 2660, P = 3.0e-241, identities = 534/535
> 10 exons

Entry AC004111 from database EMBL:
Homo sapiens BAC clone RG103H13 from 7q31, complete sequence.
Score = 598, P = 5.8e-17, identities = 128/137
1 exon

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 253 bp to 1092 bp; peptide length: 280
Category: similarity to unknown protein

1 MIEHKIVIA DVKLVADFQRI YILYWRKRFT EQPITDFCSV IRINSTAPFE
51 EQENYFLLCD VLPEDRILRE ELQKQRLREI LEQQQQRND NNFHGVCMFC
101 NEEFLGNRSV ILNHMAREHA FNIGLPDNIV NCNEFLCTLQ KKLDNLQCLY
151 CEKTFRGKNT LKDHMRKKQH RKINPKNREY DREYVINYLE LGKSWEVQL
201 EDDRELLDHQ EDDWSDWEEH PASAVCLFCE KQATIEKLY VHMDAHEFD
251 LLKIKSELGL NFYQQVKLVN FIRRVHQCR

BLASTP hits

Entry CEF46B6_6 from database TREMBLNEW:
product: "F46B6.7"; Caenorhabditis elegans cosmid F46B6
>TREMBL:CEF46B6_6 product: "F46B6.7"; Caenorhabditis elegans cosmid
F46B6
Score = 630, P = 1.1e-61, identities = 123/289, positives = 183/289

Entry AF059531_1 from database TREMBLNEW:
gene: "PRMT3"; product: "protein arginine N-methyltransferase 3"; Homo
sapiens protein arginine N-methyltransferase 3 (PRMT3) mRNA, partial
cds. >TREMBL:AF059531_1 gene: "PRMT3"; product: "protein arginine
N-methyltransferase 3"; Homo sapiens protein arginine
N-methyltransferase 3 (PRMT3) mRNA, partial cds.
Score = 120, P = 1.5e-04, identities = 23/78, positives = 42/78

Entry YB9M_YEAST from database SWISSPROT:
34.7 KD PROTEIN IN SHM1-MRPL37 INTERGENIC REGION.
Score = 112, P = 4.6e-04, identities = 43/165, positives = 71/165

Alert BLASTP hits for DKFZphfkd2_47a4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphfkd2_47a4, frame 1

Report for DKFZphfkd2_47a4.1

[LENGTH]	280
[MW]	33921.94
[pI]	5.63
[HOMOL]	TREMBL:CEF46B6_5 gene: "F46B6.7"; Caenorhabditis elegans cosmid F46B6 1e-56
[BLOCKS]	BL01032B Protein phosphatase 2C proteins
[BLOCKS]	BL00028 Zinc finger, C2H2 type, domain proteins
[PROSITE]	MYRISTYL 1
[PROSITE]	ZINC_FINGER_C2H2 1
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 3
[PROSITE]	TYR_PHOSPHO_SITE 2
[PROSITE]	PKC_PHOSPHO_SITE 2
[PROSITE]	ASN_GLYCOSYLATION 2
[PFAM]	Zinc finger, C2H2 type
[KW]	Alpha_Beta
[KW]	LOW_COMPLEXITY 8.21 %

SEQ MIEHKIVIA DVKLVADFQRI YILYWRKRFT EQPITDFCSV IRINSTAPFE EQENYFLLCD
SEG
PRD cccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeccccccchhhheeecc

SEQ VLPEDRILRE ELQKQRLREI LEQQQQRND NNFHGVCMFC NEEFLGNRSV ILNHMAREHA
SEG
PRD cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeccccccccceehhhhhhh

SEQ FNIGLPDNIV NCNEFLCTLQKKLDNLQCLYCEKTFRGKNT LKDHMRKKQH RKINPKNREY

```

SEG .....
PRD hccccccccchhhhhhhhhhhhhhhheccccccccchhhhhhhhhhhhhcccccccccc

SEG DRFYVINYLELGKSWEEVQLEDDRELLDHQEDDSDWEEHPASAVCLFCEKQAEETIEKLY
SEG .....
PRD ceeeeeeeccccchhhhhhhcchhhhhccccccccccccccccchhhhhhhhhhhhhhh

SEG VHMEDAHEFDLLKIKSELGLNFYQQVKLVNFIRRVHQR
SEG .....
PRD hhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhccccc

```

Prosite for DKF2phkd2_47a4.1

PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	107->111	ASN_GLYCOSYLATION	PDOC00001
PS00004	27->31	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	160->163	PKC_PHOSPHO_SITE	PDOC00005
PS00006	160->164	CK2_PHOSPHO_SITE	PDOC00006
PS00006	194->198	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00007	178->185	TYR_PHOSPHO_SITE	PDOC00007
PS00007	13->22	TYR_PHOSPHO_SITE	PDOC00007
PS00008	124->130	MYRISTYL	PDOC00008
PS00028	148->171	ZINC_FINGER_C2H2	PDOC00028

Pfam for DKF2phkd2_47a4.1

HMM_NAME	Zinc finger, C2H2 type		
HMM	*CpwPDCgKtFzrwsNLrRHRM..T.H*		
	C +	C+KTFR + +L+ HMR	H
Query	148	CLY--CEKTRFGKNTLKDHMRKK-QH	170

DKFZphfkd2_4b6

group: kidney derived

DKFZphfkd2_4b6 encodes a novel 133 amino acid protein with similarity to Homo sapiens clone 25003 partial CDS.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to Homo sapiens clone 25003

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1936 bp

Poly A stretch at pos. 1916, polyadenylation signal at pos. 1890

```
1 GGGAGACTTG CAATGAAGTT AGAATGAACA GGAGGAGTCT GCAGCTTTTC
51 AGTGCCTGGG ATAACATATAG TTAAAGATC ATTGTGTAAA ATAGGATTTT
101 TAGTCAGCAT GCATTGTTT AAACCGACTA ACTGATAGCC TAAAACTTTA
151 TTTTGCATT TTGCCAATCC TTGGAGTTT GTTTGCAGA ATTAAGAAAA
201 AAATGAATGT ATGATCATCT GAAAAGGGCT TTCTCTCAAT CCCACTTCAT
251 GGCATGACCT CTGCTGGATC ATTAGTTCTA GCCAGAGAAG TAGCAAAGGA
301 ACATGACCTC TGAGACCTCC CTCCCTCAT CAGTGGGGCT GACTGAGCTG
351 GGGGCTTGAA GCCGGAGSTA ACCTTCCTG TCGAATGTTT CTTTAGAGAA
401 TGGCAATGGT CTCTGCGATG TCCTGGGTCC TGTATTGTG GATAAGTGCT
451 TGTGCAATGC TACTCTGCCA TGGATCCCTT CAGCACACTT TCCAGCAGCA
501 TCACCTGCAC AGACCAGAAG GAGGGACGTG TGAAGTGATA GCAGCACACC
551 GATGTTGCAA CAAGAATCCG ATTGAGGAGC GGTCAACAAC AGTAAAGTGT
601 TCCTGTCTAC CTGGAAGAAG GGCTGGAACA ACAAGAAACC GGCCTTCTTG
651 CGTCGATGCC TCCATAGTGA TTTGAAATG GTGGTGTGAG ATGGAGCCTT
701 GCCTAGAAGG AGAAGAATGT AAGCACTCC CTGACAATTG TGGATGGATG
751 TGGCAACAG GCAACAAAAT TAAGACCACG AGAATTACCC CAAGAACCTA
801 ACAGAAGCAT TTGTGGTAGT AAAGAAAAC CAACCTCTG GAAAAACATC
851 TTTGAGAATC TCAACATCT CACATATATA CAAGCCAAAT GGAATTCCTA
901 CTTGCACCTT GACTGGCTAC CAGATAATCA CAGTGCCTT AGTGTGTGTA
951 ACCGAAATAT CTACAGTGAG AAGACACAGC GTTTTGGCAT CACCATGGAA
1001 AGTGGGCTTA AAAAAGGGTC TTCTCAGTGA AATTTTGGG CATCATGAAG
1051 AACGATCAAC TATCTTCTAA TTTGAATCTA TAGTTACTTT GTACCATTTG
1101 AAATATATGT ATATATATAT ATATAATATT TTGAAATATT ATCTATTCTC
1151 TTCAAGAAAT GAACAGTACC ACAGTTTGAG ACGGCTGGTG TACCCCTTTG
1201 AGTTTTGGAT GTTTTGTCTG TTTTGTCTTG TTTTGTAGT CATTTCTTTT
1251 TCTAACGGCA AGGAAGATAT GTGCCCTTTT GAGAATTCAA GATGGCACTG
1301 ACACGGGAAG GCCAGCTACA GGTGGACTCC TGGAAATTGA GGCATCATAA
1351 TGATACTGAA TCAAGAACTT CCTTCTGCTT CTACCAGATG GCCCAAGGAA
1401 GCACATCGTC CTGTTTTATT GCTTCTACCC CTGTGCAATA TTAGCATGCA
1451 AGCTTGGCTT ACATAGTCAT ACTTTATATT CAATTGATAT ATAATAACCG
1501 TTCTAACCTC TTCCAGGAAA ATATTTTATG AACTACTAGC TTTTCCACTT
1551 AGAAGAAAAA GAGGATTCTT AAGGGAGCCA CTCCACCATG CTATTAAGAC
1601 TCTGGCAGAG TTATGGGTAG CATATGGATC CCTACATGAA TAAGTCTCTG
1651 AATACAATG TCTTAAGGCT TTGTATAGCT GTCTAGACT GCAGAAATGT
1701 CCTCTGATTA AATCCAAACT CTGGCATCGT TAACATACATA GTGCTGTAGC
1751 AACAACTCTT ATCATGGCAT CTCCTTCTAT GTTTGTTTGT CTTTTCCTAA
1801 GAGTATTCAG GTCTCCTCTT GTGAGATAGG AAGGCCATGA AAACAATTAG
1851 ATTTCAAGAT GATCTATGTG ACCAAATGTT GGACAGCCCT ATTAAGTGG
1901 TAAACAACCT CTTTCTAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

100 90 80 70 60 50 40 30 20 10 0

```

1 MAMVSAMSWV LYLWISACAM LLCHGSLQHT FQQHHLHRPE GGTCEVIAAH
51 RCCNKNRIEE RSQTVKCSCL PGKVAGTTRN RPSCVDASIV IWKWWCEMEP
101 CLEGEECKTL PDNSGWMCAT GNKIKTTRI H PRT

```

Alert BLASTP hits for DKFZphfkd2_4b6, frame 1

.....

(No Pfam data available for DKF2phfkd2_4b6.1)

DKFZphfkd2_4c8

group: kidney derived

DKFZphfkd2_4c8 encodes a novel 153 amino acid protein with partial similarity to huntington's associated protein HAP1.

The novel protein contains a leucine zipper involved in protein-protein interaction.
No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of kidney-specific genes.

similarity to KIAA0549 and HAP1

potential frame shift at Bp ~1350-1500 will be checked

Sequenced by GBF

Locus: unknown

Insert length: 3182 bp

Poly A stretch at pos. 3162, polyadenylation signal at pos. 3135

```
1 GGGCTTCCCC CATAGAATTT TTCTTTTCAT TGCCCACTTT ACTGTTTTGG
51 CTCACAGCTG TCGTTAAGAA TGTACAGCCT AATTCTGGTG TGTTCGGGA
101 TATTCTTCTG TCCAGTATTC TGAAGGGCG GGGAGGCATG GCAGCGTTTT
151 ACTTGACGTT GATGGTGCTG TGAAGTCCAT TCTTTCCTCT GCAAGACTAC
201 TGACTATGCA GAAATTTATC GAAGCGGATT ATTATGAACT AGACTGGTAT
251 TATGAAGAAAT GCTCGGATGT TTTATGTGCT GAAAGAGTTG GCCAGATGAC
301 TAAGACATAT AATGACATAG ATGCTGTAC TCGGCTTCTT GAGGAGAAAG
351 AGCGGGATTT AGAATTGGCC GCTCGCATCG GCCAGTCGTT GTTGAAGAAG
401 AACCAAGACCC TAACCGAGAG GAACGAGCTG CTGGAGGAGC AGGTGGAACA
451 CATCAGGGAG GAGGTGTCCT AGCTCCGGCA TGAGCTGTCC ATGAAGGATG
501 AGCTGCTTCA GTTCTACACC AGCGCAGCGG AGGAGAGTGA GCCCGAGTCC
551 GTTTGCTCAA CCCCCTTGAA GAGGAATGAG TCGTCCTCCT CAGTCCAGAA
601 TTACTTTTCA TTGATTCTCT TTCAAAAGAA GCTGAAAGAC CTTGAAGAGG
651 AGAATGTTGT ACTTCGATCC GAGGCCAGCC AGCTGAAGAC AGAGACCATC
701 ACCTATGAGG AGAAGGAGCA GCAGCTGGTC AATGACTGCG TGAAGGAGCT
751 GAGGGATGCC AATGTCCAGA TTGCTAGTAT CTCAGAGGAA CTGGCCAAGA
801 AGACGGAAGA TGCTGCCCGC CAGCAAGAGG AGATCACACA CCTGCTATCG
851 CAAATAGTTG ATTTGCAGAA AAAGGCAAAA GCTTGCCGAG TGGAAAATGA
901 AGAAGTTGCT CAGCATCTGG GGGCTGTCAA GGATGCCAGC CGGCAGCTCA
951 CAGCCGAGCT GCGTGAGCTG GAGGACAAGT ACGCAGAGTG CATGGAGATG
1001 CTGCATGAGG CGCAGGAGGA GCTGAAGAAC CTCCGGAACA AAACCATGCC
1051 CAATACCACG TCTCGGCGCT ACCACTCACT GGGCCTGTTT CCTATGGATT
1101 CCTTGGCAGC AGAGATTGAG GGAACGATGC GCAAGGAGCT GCAGTTGGAA
1151 GAGGCCGAGT CTCCAGACAT CACTCACCAG AAGCGTGTCT TTGAGACAGT
1201 AAGAAACATC AACCAGGTTG TCAAGCAGAG ATCTCTGACC CCTTCTCCCA
1251 TGAACATCCC CGGCTCCAAC CAGTCTCGG CCATGAACTC CCTCCTGTCC
1301 AGCTGCGTCA GCACCCCGCG GTCCAGCTTC TACGGCAGCG ACATAGGCAA
1351 CGTCTCTCTC GACAACAAGA CCAACAGCAT CATTCTGAA ACAGAGGCAG
1401 CCGACCTGGG AACAGATGAG CGGAGTAAGA AGCCGGGGAC GCCGGGCACC
1451 CCCAGGCTCC CACGACCTGG AGACGGGCGT GAGCGGGCTG TCCCTGGGCC
1501 GGGAGAACTA CCTCTCGGAG AGGAGGTTCT TTGAGGAGGA GCAAGAGAGG
1551 AAGCTCCAGG AGCTGGCGGA GAAGGGCGAG CTGGCGAGCG GCTCCCTCAC
1601 ACCCACTGAG AGCATCATGT CCCTGGGCAC GCACCTCCCG TTCTCCGAGT
1651 TCACCGGCTT CTCTGGCATG TCCTTCAGCA GCCGCTCCTA CCTGCCTGAG
1701 AAGCTCCAGA TCGTGAAGCC GCTGGAAGGT GATCACGCGG GGCCTCGGCC
1751 CCTCTCTGTC CTCCTGGGGG ACTCCCTTTG GTCCCTGATC CACCTGCGGA
1801 AGGGGGGGCA CCTCTGTAC GCCTACTCCT TTTTCTTCG CGACAGCCAC
1851 CCGCGCTGCT GGTTTGAGTT CCTCTGAGGG TGGTGCTCAG CCTAGGCCCTC
1901 CGTCCCTCCC CTCTGGCTGG CAGGTGTGAC AATGCACACA TAGGCCATGA
1951 AACTCGCCGA GGAAAGACAA GCATGTGCAC TGTGGTCTTC TAGTTCTTTC
2001 CTTTGCCTTT AGAACCTTAG AAATAAAAC TTTTGTGGCG GTAGAGGCAC
2051 TGCTAACTGA TTCAAAAATT AATTAGGTTT TGCCGTGGGG TGTGAGGAAT
2101 GCAGAAAATT AATGCTTTAG CTTTTCTGCA GTTTTGGTGT CGGGGAGAGG
2151 TTCCAAAGCA ACTCTATTAA ATGGGGATTT TTTTTCCTCC ATAACCACCT
2201 GAATGTGATT TGTGGGCTTA TGTGTTCTGA TTTGAACCTC ATATAGCAAG
2251 GTTGTGGCTT TTGGCAGATG CAGTATGTTG TGAGCGCGCG TCCTAGAGTC
2301 TACAATTGAG AGTCCAGGAA GGGGTGGCTG TGGAGACAAG TGAGTTTGTG
2351 ACCTCCGTAA GCCACCCTTT TTCAGGCTCA GTTCATGTGT TAGTATCAGG
2401 GGCATCTCAG ATGATTAAAC TCATGGGAAA AACTTCTCTC TTCCCTCTCT
2451 CCCTCTTGCC CTCCTGCCTC TTTTCTTTT TTTTCTTTT AATTGGGCA
2501 CTTATAAAAT GTTTTCCCTC TACCTGCTGC TACTCTGCCA AGAGCCACCA
2551 AGTGCTTATA TTTTTCATTT TTTACTCCTT TAGTTGGAA AGCCATATAC
2601 GTTTGAGAAG GTGTTTTAAA ACTCTGTGTT ACACCTACGA TGCAAGGCA
2651 AATCAGAACT TCTGTAAGGC AGAACTTTCC CAACCTTAAA AAAATTATTG
```



```

2701 TCCCTCTAG GAGCCTTCTT AGACGTTTT TCCTAATCAC CCCCCAAAGA
2751 CATTTAATA CCACATATAT ATTGTTTATG TACTATATGT ATATACATAA
2801 ACAATACATA AGCAATACAT CTGTGGTATT AAAATTAAAA AGAATCCAAT
2851 TATGTTTACC TCAAAGAAGC CTGTTTTTGC TTCTTGGGAG CAATATTGCC
2901 CCTGTGAGAC TGCATGCTAT AAGGTAAGGT TGTGCTTGT AAAGACCCAA
2951 GACATGACTG GGTTCACAG TCTCAAAGG AAGAGGGTGG GCTAGTTGT
3001 TTTTATTATT ATTTTAAAT TGTATAATG GGCTCTTCT TAGAGTTTCT
3051 AAAAGGTATA GCTTACTCTT TTTTAATTGT TTATTAGTT GTAAGCTTAG
3101 TGATTGTTT CTGATCCACA TTGTGTGTGT TCTTCAATAA AATCTTTCAT
3151 TTCTGCAATT TTAAAAAAA AAAAAAAA AA

```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 206 bp to 1531 bp; peptide length: 442
 Category: similarity to known protein
 Classification: unset
 Prosite motifs: LEUCINE_ZIPPER (139-161)

```

1 MQKFIADYY ELDWYEECS DVLCAERVGQ MTKTYNDIDA VTRLLEEKER
51 DLELAARIGQ SLLKKNKTLT ERNELLEEQV EHIREEVSQL RHELSMKDEL
101 LQFYTSAAEE SEPESVCSTP LKRNESSSV QNYFHLDSIQ KKLKDLLEEN
151 VVLRSEASQL KTETITYEEK EQQLVNDVCV ELRDANVQIA SISEELAKKT
201 EDARQEQEI THLLSQIVDL QKKAKACAVE NEELVQHLGA AKDAQRLTA
251 ELRELEDKYA ECMEMLEHAQ EELKRLNKTP MPNTTSRRYH SLGLFPMDSL
301 AAEIEGTRMK ELQLEEAESP DITHQKRVFE TVRNINQVVK QRSLTPSPMN
351 IPGSNQSSAM NSLLSSCVST PRSSFYGSID GNVVLNKTN SIILETEAAD
401 LGNDERSKKP GTPGTPRLPR PGDGAEEAVP APGELPLGEE VL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phkd2_4c8, frame 2

PIR:S72555 huntingtin-associated protein HAP1 - human (fragment), N = 1, Score = 234, P = 8.6e-19

TREMBL:CEUT27A3 7 gene: "T27A3.1"; Caenorhabditis elegans cosmid T27A3., N = 1, Score = 226, P = 9.9e-16

PIR:S67495 huntingtin-associated protein HAP1-A - rat, N = 1, Score = 215, P = 1.6e-14

>PIR:S72555 huntingtin-associated protein HAP1 - human (fragment)
 Length = 320

HSPs:

Score = 234 (35.1 bits), Expect = 8.6e-19, P = 8.6e-19
 Identities = 66/189 (34%), Positives = 110/189 (58%)

Query: 109 ESEPEPESVCSTPLKRNE--SSSSVQNYFH---LQSLQKKLKDLEENVVLRSEASQLKTE 163
 EE+E + C+ P + S ++ + H L++LQ+KL+ LEEEN LR EASQL T

Sbjct: 28 EEAEEDLQCAHPCDAPKLISQEALLHQHHCPOLEALQEKRLLEEENHQLREEASQLDT- 86

Query: 164 TITYEEKEQQLVNDVCVKELRDANVQIASISEELAKKTEDAARQEEITHLLSQIVDLQKK 223

E++EQ L+ +CV++ +A+ Q+A +SE L + E+ RQQ+E+ L +Q++ LQ++
 Sbjct: 87 ---LEDEEQMLILECVEQFSEASQMAELSEVLVLRLENYERQQQEVARLQAVLKLQQR 143

Query: 224 AKACAVENEELVQHLGAAKDAQRLTAEE--LRELEDKYAECE--MLHEAQEELKNL-RN 278
 + E E+L + L + K+ Q QL E L ++ AE + + + + + RN

Sbjct: 144 CRMYGAETKLOKQLASEKEIQMOLQEEETLPGFQETLAEELRTSLRRMISDPVYFMERN 203
Query: 279 KTMP--NTTSRRY 289
MP +T+S RY
Sbjct: 204 YEMPRGDTSSLRY 216

Peptide information for frame 3

ORF from 1416 bp to 1874 bp; peptide length: 153
Category: similarity to known protein
Classification: unset

1 MSGVRSRGRR APPGSHDLET ALRRLSLRRE NYLSERRFFE EEQERKLQEL
51 AEKGELRSGS LPTESIMSL GTHSRFSEFT GFSGMSFSSR SYLPEKLQIV
101 KPLEGDHAGP RPLSVLLGDS LWSLIHLRKA GHLCCHAYSFF FRDSHPRCWF
151 EFL

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphkd2_4c8, frame 3

TREMBL:AB011121.1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds., N = 1, Score = 252, P = 5.5e-21

>TREMBL:AB011121.1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens mRNA for KIAA0549 protein, partial cds.
Length = 469

HSPs:

Score = 252 (37.8 bits), Expect = 5.5e-21, P = 5.5e-21
Identities = 57/98 (58%), Positives = 69/98 (70%)

Query: 8 GRRAPPGSHDLETALRRLSLRRENYLSERRFFEEEQERKLQELAEXGELRSGSLPTESI 67
G+ P G DL TAL RLSLRR+NYLSE++FF EE +RK+Q LA++ E SG +TPTES+
Sbjct: 27 GQPGPSGDSLDLALHRLSLRRQNYLSEKQFFAEWQRKIQLADQKEGVSGCVTPTESL 86
Query: 68 MSLGTHSRFSEFTGFSGMSFSSRSYLPEKLQIVKPLEG 105
SL T SE T S S R ++PEKLQIVKPLEG
Sbjct: 87 ASLCTTQ--SEITDLSSAS-CLRGFMPEKLQIVKPLEG 121

Pedant information for DKFZphkd2_4c8, frame 2

Report for DKFZphkd2_4c8.2

{LENGTH} 442
{MW} 50020.14
{pI} 4.77
{HOMOL} TREMBL:AF040723_1 product: "neuroan1"; Homo sapiens neuroan1 mRNA, complete
cds. 5e-29
{FUNCAT} 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
5e-08
{FUNCAT} 30.04 organization of cytoskeleton [S. cerevisiae, YIL149c] 5e-08
{FUNCAT} 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-08
{FUNCAT} 03.04 budding, cell polarity and filament formation [S. cerevisiae, YIL138c]
6e-08
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YGR130c] 2e-07
{FUNCAT} 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-06
{FUNCAT} 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-06
{FUNCAT} 1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1643] 1e-06
{FUNCAT} 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 -
myosin-1 isoform] 3e-06
{FUNCAT} 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-06
{FUNCAT} 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YKR095w] 4e-06
{FUNCAT} 30.10 nuclear organization [S. cerevisiae, YKR095w] 4e-06
{FUNCAT} 03.13 meiosis [S. cerevisiae, YNL250w] 2e-05
{FUNCAT} 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 2e-05

[FUNCAT] 08.99 other intracellular-transport activities (S. cerevisiae, YNL079c)
5e-05
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YNL079c] 5e-05
[FUNCAT] 10.05.99 other pheromone response activities (S. cerevisiae, YHR158c)
1e-04
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 1e-04
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YNL272c] 3e-04
[FUNCAT] 08.16 extracellular transport (S. cerevisiae, YNL272c) 3e-04
[BLOCKS] BL01289B
[BLOCKS] BL00415M Synapsins proteins
[EC] 3.6.1.32 Myosin ATPase 2e-07
[PIRKW] tandem repeat 2e-07
[PIRKW] heterodimer 1e-06
[PIRKW] endocytosis 9e-07
[PIRKW] heart 1e-06
[PIRKW] transmembrane protein 4e-07
[PIRKW] zinc finger 9e-07
[PIRKW] metal binding 9e-07
[PIRKW] DNA binding 3e-06
[PIRKW] muscle contraction 2e-07
[PIRKW] acetylated amino end 3e-06
[PIRKW] actin binding 2e-07
[PIRKW] mitosis 1e-06
[PIRKW] microtubule binding 1e-06
[PIRKW] ATP 2e-07
[PIRKW] chromosomal protein 1e-06
[PIRKW] receptor 3e-08
[PIRKW] thick filament 2e-07
[PIRKW] phosphoprotein 8e-06
[PIRKW] glycoprotein 3e-08
[PIRKW] skeletal muscle 3e-06
[PIRKW] DNA condensation 1e-06
[PIRKW] alternative splicing 2e-06
[PIRKW] coiled coil 2e-07
[PIRKW] P-loop 2e-07
[PIRKW] heptad repeat 4e-07
[PIRKW] methylated amino acid 2e-07
[PIRKW] peripheral membrane protein 9e-07
[PIRKW] cardiac muscle 6e-06
[PIRKW] hydrolase 2e-07
[PIRKW] muscle 2e-06
[PIRKW] cytoskeleton 2e-06
[PIRKW] Golgi apparatus 4e-07
[PIRKW] calmodulin binding 9e-07
[SUPFAM] myosin motor domain homology 2e-07
[SUPFAM] tropomyosin TPM1 2e-06
[SUPFAM] giantin 4e-07
[SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
[SUPFAM] human early endosome antigen 1 9e-07
[SUPFAM] unassigned kinesin-related proteins 4e-07
[SUPFAM] M5 protein 8e-08
[SUPFAM] cytoskeletal keratin 3e-06
[SUPFAM] myosin heavy chain 2e-07
[SUPFAM] conserved hypothetical P115 protein 1e-06
[SUPFAM] centromere protein E 1e-06
[SUPFAM] pleckstrin repeat homology 2e-06
[SUPFAM] kinesin motor domain homology 4e-07
[PROSITE] LEUCINE_ZIPPER 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 6.79 %
[KW] COILED_COIL 27.15 %

SEQ MQKFIEADYYELDWYEECSVDLCAERVGQMTKYNDIDAVTRLLEEKERDLELAARIGO
SEGXXXXXXXXXXXXXXXXXXXX
PRD ccc
COILSC

SEQ SLLKKNKTLTERNELLEQVEHIREEVSQLRHELMSKDELQFYTSAAEESEPEVCSTP
SEG
PRD hhh
COILS CC.....

SEQ LKRNESSSVQNYFHLDSLQKKLKDLEENVVLRSEASQLKTETITYEEKEQQLVNDVCV
SEG
PRD hhh
COILSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

```
SEQ      ELRDANVQIASISEELAKKTEDAARQQEEITHLLSQIVDLQKKAKACAVENEELVQHLGA
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    ..CCCCCCCCC...

SEQ      AKDAQQLTAELELEDKYAECMEMLHEAQEELKNLRNKTMTNPNTTSRRYYSIGLFFMDSL
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ      AAIEEGTMRKELQLLEEASPOITHQKRVFETVRNINQVVQRSLTPSPMNI PGNSOSSAM
SEG      .
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccchhhh
COILS    .....

SEQ      NSLLSCSVSTPRSFSYFGSDIGNVLVDNKNTSIILETEAADLGNDERSKKPGTPTGPLRPR
SEG      .xxxxxxxxxxxxx
PRD      hhhhccccccccccccccccceeeccccceccccccccccccccccccccccccccc
COILS    .....

SEQ      PGDGAAEAAPAGPLPLGEEVL
SEG      xxxxx
PRD      ccccccccccccccccccccccc
COILS    .....
```

Prosite for DKF2phfkd2 4c8.2

PS00029 139->161 LEUCINE ZIPPER PDOC00029

(No Pfam data available for DKFZphfkd2_4c8.2)

Pedant information for DKFZphfk2:4c8, frame 3

Report for DKFZphfkd2 4c8.3

```
{LENGTH}      153
[MW]            17642.03
[pI]            9.38
[HOMOL]         TREMBL:AB011121_1 gene: "KIAA0549"; product: "KIAA0549 protein"; Homo sapiens
mRNA for KIAA0549 protein, partial cds. 2e-12
[KW]            Alpha_Beta
[KW]            LOW_COMPLEXITY      12.42 %
```

[illegible]

(No Prosite data available for DKF2phfk2 4c8.3)

(No Pfam data available for DKFZphfkd2 4c8.3)

DKFZphfkd2_4k14

group: intracellular transport and trafficking

DKFZphfkd2_4k14.3 encodes a novel 254 amino acid putative GTP-binding protein nearly identical to Rab6.

Rab proteins are members of the Ras superfamily of GTPases. Rab proteins are localised to the cytoplasmic side of organelles and vesicles involved in the secretory (biosynthetic) and endocytotic pathways in eukaryotic cells. Rab proteins direct the targeting and fusion of transport vesicles to their acceptor membranes.
rab6 is a ubiquitous ras-like GTPase involved in intra-Golgi transport.

The new protein can find application in modulating the transport of vesicles inside the Golgi apparatus.

strong similarity to Rab6

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3084 bp

Poly A stretch at pos. 3061, polyadenylation signal at pos. 3043

```
1 GGGGCACTCA GCAGGTTGGG CTGCGGCGGC GGGGCTGGG GAAGCCGAAG
51 CGCCGCGCGT GAGAGATCCC GGATACATCT CGGTTTGGG CTCCGCCACC
101 CTCCGTCTCT CTCCGCGAGG TCTCTGAGCC GGGTGGGAA GGAGGGAACG
151 GCGCTAGCCT TGGGAAGCCA AAGCACACCC CTGGTCCCG CCGACACCCG
201 CCTCCTTCCC TTCCAGCCGG CGGGCTCGC TCCGTGCTCG GCTACTCTCG
251 CGGGAGGCGG CGGGGCTGCG CAGTCTGTGG CGAGCCCTGC TGCCCTCCAG
301 CCGGCTTCTT CCAGCCGGGC TCCTCCACCG GCCCTTGCG GGGCACAGAG
351 AGCTCGGCGC CGGCCCTTCC GCTCGCCTTT TTCCTCAGCC GGCTGGAGGA
401 GCATCGGTCC GGGAGGTCTC TGSGCTGAGG CGGCGACAGC TCCTCTAGTT
451 CCACCATGTC CGCGGGCGGA GACTTCGGGA ATCCGCTGAG GAAATTCAG
501 CTGCTGTTCC TGGGGGAGCA AAGCGTTGCA AAGCATCTT TGATCACCAG
551 ATTCAAGTAT GACAGTTTGG ACAACACCTA TCAGGCAATA ATTGGCATTG
601 ACTTTTATC AAAAATATG TACTTGGAGG ATGGAACAAT CGGGCTTCGG
651 CTGTGGGATA CGGCGGGTCA GGAACGTCTC CGTAGCCTCA TTCCCAGGTA
701 CATCCGTGAT TCTGTGTCAG CTGTAGTAGT TTACGATATC ACAATGTGTA
751 ACTCATTCCA GCAAACTACA AAGTGGATTG ATGATGTCAG AACAGAAAGA
801 GGAAGTGATG TTATCATCAC GCTAGTAGGA AATAGAACAG ATCTTGCTGA
851 CAAGAGGCAA GTGTCAAGTG AGGAGGGAGA GAGGAAAGCC AAAGGGCTGA
901 ATGTTACGTT TATTGAAACT AGGGCAAAAA CTGGATACAA TGTAAAGCAG
951 CTCTTTCGAC GTGTAGCAGC AGCTTTGCCG GGAATGGAAA GCACACAGGA
1001 CGGAAGCAGA GAAGACATGA GTGACATAAA ACTGGAAAAG CCTCAGGAGC
1051 AAACAGTCAG CGAAGGGGGT TGTTCCTGCT ACTCTCCCAT GTCATCTTCA
1101 ACCCTTCCCT AGAAGCCCCC TTACTCTTTC ATTGACTGCA GTGTGAATAT
1151 TGGCTTGAAC CTTTCCCTT CATTAAATAAC GTTTTGAAT TCATCATTGC
1201 TGGCTGTCTC GTGGAGGTGA TCTATTAGCT TCACAAGCAC AAAAAAAGTC
1251 AGCGTCTTCA TTATTATAT TTTACAAAAA GCCAAATTAT TTCAGCATAT
1301 TCGGGTGATA ACTTTAAAAA TTAGATACAT TTTCTTAACA TTTTCTTCTT
1351 TTTTAATGTT ATGATAATGT ACTTCAAAAT GATGGAAATC TCAACAGTAT
1401 GAGTATGGCT TGGTAAACGA GCAGTATGTT CACAGCCTGC TTTATCTCTC
1451 CTGCTCTTTC TCACCTCTCC CTTACCCCGT TCCCTATTTC CGTGTCTTFA
1501 CCTAGCCTCC CCCCCTTCC TCAAAACAAA CAAGAGATGG CAAGCAGCA
1551 GTCCGACCAA GCCCACTGGA ATTATCCTTT AATTTTACAG ATACCACTTG
1601 CTGTAGGCTG TGGACCAAGA TGTCAGAAAT TATTCTTGAG CACTGATGTA
1651 AATTACTTAG ATCTCTTTG AGGTCAGAAT TCAGCATCA CGGTAGGCAG
1701 TGCTTGAATG AGAAAAGCCT CCGTGTGCAT CTTCAAAATG AGTCCTAAAG
1751 AACATACTGA GTACTTATAA GTAGCAGAAC ATAAATGTA TTTCTGACTA
1801 ACACAAATGG TCCTTTCACA TGTGCTTAT TAGACTCTGG GAGAGAAAAG
1851 TAACCAAGTG CTTCAGAACA GGTTTTATAG ATTTACTTCT TCATGTAAG
1901 ATAAAGAGT TCTAATGAAC TATTTCTCCC AAGGTTTAA AATTGTCAAG
1951 AGTTATTCTG TTTGTTTAAA AAGTAAGAAA CCTCTGTAAG CAATAGATTT
2001 TGCTTGGGTT TTCTTCTTA AAAAAATAAT ACTATGCAGG CAAGACACCA
2051 TAAAGTTTAA ATCTCTTACA GAAGAACCAG TGAAGCAATT TAAATTTGGC
2101 ACTACGATCA AACTACTGTA ATTAGCAGAA ATAACGATAT CTAAGCTTA
2151 CCAGCAAAAG AACCTCAGC AGAATAGCAA AAACCTTGCT CAGGACATTT
2201 GAGGTCAAAAT TGAAGACGGA AGACGGAAAC CGGAAACCGT TTTCTGTGAA
2251 GCCCTAGAG GCAGATCAGG TAAGCATACA TAGTAGAGGG AAAGGAGAGA
2301 ATGGAATAAA AACTGAATAT TATGCAGATT TATGCCTTAT TTTTATGAT
2351 TTTTAAAGGT TGGGTCTTTC AGGCTGGTTT TGGTTTGAT TAGATCTGTA
2401 TAGTTTAGTG ATTTAGTTTT ATATTTAAGC TACGATTAAT ATTTTCTT
2451 TGGCGATATT TCTTGCTTT TTTTAAAAA CAACTTTCCA TTTTATAGATG
```

```

2501 TTTGCTTGAA TCTATTTAGA GCTTCACCAT GGCAATATGT ATTTCCCTTA
2551 AAACACTGCA AACAAATATA CTAGGAGTGT GCCCTTTTAA TCTTTACTAG
2601 TTATTGTGAG ACTGCTGTGT AAGCTAATAA ACACATTGTG AAAAACATTG
2651 TTTGCAGGAA GAAAACTTCG AGTTACAGGT CAGGAAAGC CTGCTGAATT
2701 TATGTTGTAA ACGTTACTTA ACACAGTATA AAGATGAAAA GACAACAAAA
2751 GTATCTTCAT ACTTCCTCAT CCCCTCATTG CAACAAACC TTAACCTGGG
2801 AGAACCTTAG TCCCCTCTCT TTCTCTTCC TCCTCCACTT CCCACTTATT
2851 GCCACTTGTG AATATTGAGA GAGCACTTGG ATTATGGATC TGAATAGAGA
2901 AATGCTTACA GATAATCATT AGCCACATA CCAGTAACCT ATACTTAAAG
2951 ATGGGATGGA GTTATAAAGT GCTTTTATAA TCCATATATA TTGCTAAAGG
3001 CAAGGGTTGA CTCITTTGTT TATTTTGACA TGGCATGTCC TGAATAAAT
3051 ATGGTTTAC TATGAAAAA AAAAAAAAAA AAAA

```

BLAST Results

No BLAST result

Medline entries

98382468:

Rab proteins.

97203146:

GTP-bound forms of rab6 induce the redistribution of Golgi proteins into the endoplasmic reticulum.

Peptide information for frame 3

ORF from 456 bp to 1217 bp: peptide length: 254

Category: strong similarity to known protein

Classification: unset

Prosites motifs: BACTERIAL_OPSIN_RET (45-57)

```

1 MSAGGDFGNP LRFKLVFLG EQSVAKTSLI TRFRYDSFDN TYQAIIGIDF
51 LSKTMYLEDG TIGLRLWDTA GQERLRLSLP RYIRDSAAAV VVYDITNVNS
101 FQOTTKWIDD VRTERGSDVI ITLVGNRTDL ADKRQVSVEE GERKAKGLNV
151 TFIETRAKTG YNVKQLFRRV AAALPGMEST QDGSREDMSD IKLEKPQEQT
201 VSEGGCSCYS PMSSSTLPQK PPYSFIDCSV NIGLNLFPPL ITFCNSSLLP
251 VSWR

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphkd2_4k14, frame 3

PIR:G34323 GTP-binding protein Rab6 - human, N = 1, Score = 944, P = 6.5e-95

TREMBL:CET25G12_2 gene: "T25G12.4"; Caenorhabditis elegans cosmid T25G12., N = 1, Score = 756, P = 5.4e-75

TREMBL:NTNTRAF_1 gene: "Nt-rab6"; Nicotiana tabacum SR1 Nt-rab6 mRNA, complete cds., N = 1, Score = 698, P = 7.6e-69

TREMBL:D84314.1 product: "rab6"; Drosophila melanogaster mRNA for rab6, complete cds., N = 1, Score = 836, P = 1.9e-83

PIR:T01588 small GTP-binding protein F16B22.10 - Arabidopsis thaliana, N = 1, Score = 704, P = 1.8e-69

>PIR:G34323 GTP-binding protein Rab6 - human
Length = 208

HSPs:

Score = 944 (141.6 bits), Expect = 6.5e-95, P = 6.5e-95
Identities = 186/208 (89%), Positives = 190/208 (91%)

Query: 1 MSAGGDFGNPLRKFKLVFLGEQSVAKTSLITRFYDSFDNTYQAIIGIDFLSKTMYLEDG 60
MS GGDGDFGNPLRKFKLVFLGEQSV KTSLITRF YDSFDNTYQA IGIDFLSKTMYLED
Sbjct: 1 MSTGGDFGNPLRKFKLVFLGEQSVGKTSLITRFYDSFDNTYQATIGIDFLSKTMYLED 60

Query: 61 TIGRLRLWDTAGQERLRSIPRYIRDSAAAVVYDITNVNSFQQTWKIDDVTERGSDVI 120
T+ L+LWDTAGQER RSLIP YIRDS AVVVYDITNVNSFQQTWKIDDVTERGSDVI
Sbjct: 61 TVRLQLWDTAGQERFRSLIPSYIRDSAVVVYDITNVNSFQQTWKIDDVTERGSDVI 120

Query: 121 ITLVGNRTDLADKQVSVEEGERKAKGLNVTFIETRAKTGVNVKQLFRVAAALPGMEST 180
I LVGN+TDLADKQVSV+EEGERKAK LNV FIET AK GYNVVKQLFRVAAALPGMEST
Sbjct: 121 IMLVGNKTDLADKQVSVIEEGERKAKELNVMFIETSAKAGYNVVKQLFRVAAALPGMEST 180

Query: 181 QDGSREDMSDIKLEKPEQTVSEGGCSC 208
QD SREDM DIKLEKPEQ VSEGGCSC
Sbjct: 181 QDRSREDMDIKLEKPEQPVSEGGCSC 208

Pedant information for DKFZphkd2_4k14, frame 3

Report for DKFZphkd2_4k14.3

[LENGTH] 254
[MW] 28385.29
[pI] 7.58
[HOMOL] PIR:G34323 GTP-binding protein Rab6 - human 1e-102
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YLR262c]
7e-60
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YLR262c] 7e-60
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YOR089c] 2e-33
[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR089c] 2e-33
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YOR089c] 2e-33
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YOR089c]
2e-33
[FUNCAT] 09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 3e-28
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 8e-27
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w]
8e-27
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YOR101w]
2e-21
[FUNCAT] 11.10 cell death [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 11.03.13 regulation of nucleotide metabolism [S. cerevisiae, YOR101w]
2e-21
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 03.99 other cell growth, cell division and dna synthesis activities [S.
cerevisiae, YOR101w] 2e-21
[FUNCAT] 10.04.07 g-proteins [S. cerevisiae, YOR101w] 2e-21
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 11.01 stress response [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YNL098c] 6e-19
[FUNCAT] 04.07 rna transport [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YOR185c] 6e-16
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 4e-13
[FUNCAT] 10.02.07 g-proteins [S. cerevisiae, YPR165w] 4e-13
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 2e-09
[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-08
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-08
[FUNCAT] 03.01 cell growth [S. cerevisiae, YNL180c] 1e-05
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR094w] 5e-05
[BLOCKS] BL01115A GTP-binding nuclear protein ran proteins
[SCOP] dlas3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domai 1e-32
[SCOP] dlml_ 3.29.1.4.2 Rac1 (Human (Homo sapiens)) 2e-51
[SCOP] d5p21_ 3.29.1.4.1 CH-p21 Ras protein (human (Homo sapiens)) 7e-53
[SCOP] dihura_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) (human (Hom 1e-46
[SCOP] dls2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) (Do 6e-60
[PIRKW] nucleus 2e-14
[PIRKW] cell cycle control 5e-15
[PIRKW] membrane trafficking 3e-71
[PIRKW] endoplasmic reticulum 1e-29
[PIRKW] phosphoprotein 1e-29
[PIRKW] prenylated cysteine 2e-36
[PIRKW] signal transduction 5e-15
[PIRKW] transforming protein 5e-30
[PIRKW] purine nucleotide binding 1e-28
[PIRKW] alternative splicing 1e-18
[PIRKW] P-loop 3e-71

```

[PIRKW]      lipoprotein 2e-36
[PIRKW]      proto-oncogene 1e-20
[PIRKW]      methylated carboxyl end 1e-20
[PIRKW]      membrane protein 1e-29
[PIRKW]      GTP binding 3e-71
[PIRKW]      thiolester bond 1e-29
[PIRKW]      Golgi apparatus 1e-29
[SUPFAM]     ras transforming protein 1e-76
[PROSITE]    BACTERIAL_OPSIN_RET 1
[PFAM]       Ras family (contains ATP/GTP binding P-loop)
[KW]         Alpha_Beta
[KW]         3D

```

```

SEQ      MSAGGDFGNPLRKFLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDG
lkao-    .....CCEEEEEEECTTTTCHHHHHHHHHHCCCCCCTTTTC-EEEEEEEEETTE

SEQ      TIGLRLWDTAGQERLRLSLIPRYIRDSAAAVVVYDITNVNSFQQTWKIDDVTERGSDVI
lkao-    EEEEEEECTTTTCHHHHHHHHHHCCCCCEEEETTTTHHHHHHHHHHHHHHHHTTCCC

SEQ      ITLVGNRTDLADKRQVSVEEGERKAKGLNVTFIETRAKTGYNVKQLFRRVAAALPGMEST
lkao-    EEEEEETTTTGGGCCCHHHHHHHHHHCCCCCEETTTTHHHHHHHHHHH.....

SEQ      QDGSREDMSDIKLEKPQEQTVSEGGCSCYSPMSSSTLPQKPPYSFIDCSVNIGLNLFPSL
lkao-    .....

SEQ      ITFCNSSLLPVSWR
lkao-    .....

```

Prosite for DKFZphkd2_4k14.3

PS00327 45->57 BACTERIAL_OPSIN_RET PDOC00291

Pfam for DKFZphkd2_4k14.3

```

HMM_NAME      Ras family (contains ATP/GTP binding P-loop)
HMM            *KLVLIQDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGktIK
               KLV++C+ +V K++L RF +++F++ Y + IG+DF++KT+++++ TI
Query          15 KLVFLGEQSVAKTSLITRFRYDSFDNTYQAIIGIDFLSKTMYLEDGTIG 63

HMM            LQINDTAGQERYRaMRPMYYRGAMGFMLVYDITNRqSFENiRnWweEIrR
               L +WDTAGQER RS+ P Y+R++ ++++VYDITN SF+ ++W++++R+
Query          64 LRLWDTAGQERLRLSLIPRYIRDSAAAVVVYDITNVNSFQQTWKIDDVRT 113

HMM            HCDrDENVPIMLVGNKCDEQQRQVStEEGQeFAREWGAIPFMETSAKTN
               + ++V+I LVGN +DL+D+RQVS EEG+ A+ ++ + F+ET AKT+
Query          114 ERG--SDVIITLVGNRTDLADKRQVSVEEGERKAKGLN-VTFIETRAKTG 160

HMM            iNVEEAfMEIvReIlqrMqe.q.NqteNinidOpsrnrk....rCCCIM*
               +NV++ F +++ +++ + ++++++I+ ++++ + +C+ +
Query          161 YNVKQLFRRVAAALPGMESTQDGSREDMSDIKLEKPQEQTVSEGGCS-C 208

```


DKFZphfkd2_4ml1

group: transmembrane protein

DKFZphfbr2-4ml1 encodes a novel 159 amino acid protein with weak similarity to the putative membrane protein YMR034c of *S. cerevisiae*.The novel protein contains 4 transmembrane regions.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of kidney-specific genes and as a new marker of neuronal cells.

weak similarity to YMR034c

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1749 bp

Poly A stretch at pos. 1727, polyadenylation signal at pos. 1713

```
1 GGGGTCCTCA AAGCCGCGG AGCAACCCCG AGGTCTTTAC TTTACAATCG
51 GCAATTGAC TTGCTCTGCT GCATGTCTGG AGGGACCAAG GAAAGTGTGG
101 AGACGCTCCA AGGATTAGGT GATCGGAGCT TGAAGAGAAA AAAAGCCAAA
151 CAAATAAACA AAACCCACCC ACCCTAACGA ATATGAGGCT GCTGGAGAGA
201 ATGAGGAAAG ACTGGTTCAT GGTGGAATA GTGCTGGCGA TCGCTGGAGC
251 TAAACTGGAG CCGTCCATAG GGTGGAATGG GGGACCACTG AAGCCAGAAA
301 TAACTGTATC CTACATTGCT GTTGCAACAA TATTCTTTAA CAGTGGACTA
351 TCATTGAAAA CAGAGGAGCT GACCACTGCT TTGGTGCATC TAAACTGCA
401 TCTTTTATT CAGATCTTTA CTCTTGCAAT CTCCCGACA ACAATATGGC
451 TTTTCTTCA GCTTTTATCA ATCACACCCA TCAACGAATG GCTTTTAAAA
501 GGTTCGACAG CAGTAGTTGG CATGCTCTCG CCGTGTCTTT CTGCAGTGAT
551 TTTAACCAAG GCAGTGGTGG GAAATGAGGC AGCTGCAATA TTTAATTGAG
601 CCTTTGGAAG TTTTGTGGTA AGTAAACATA GTTAACTTGG TCTATTACAA
651 CTTTGTCTGT GATATTGTGT ATATGAAAGA TTTAGTGAAG GCTGGATTGG
701 TTTTACTCTT TGGTTAAGTA TAAAAATTGT TGAATCTTT CATGTGCCAG
751 TATCCATACC CTGAGAGAAA GTAGTTAATG AATAAAGCAA ATGTCTCTCT
801 ACAATATATT TTGGAGGTTT GGATTTTAAA ATTCCATTTA ATGAATTCAA
851 GGAATCAATT AAAACACTAT GTGTCTCTCT ATAGAGGTTA TGTCAATATA
901 TTGATCATT AATGAGGTCT TTTAGATTAT TATTATTTTG TATCATGGGA
951 CTGAGGATT TGAAGAGGAA ACATGACCCA GCTGGTCAGA AAGGGAATGC
1001 TAATTTACTT GTTGACATGC CATTATTTT GTACATTTC CTGTCAAAGA
1051 AGCTACTGGC TTGGATGCTT CTGAGAAATC TATGTGAGAA AAAATTTGAA
1101 AGGAAGATAT GACTAATGAG TAATTTGCAA GTAAATGTTG TATCTATATA
1151 TATATATATA TAAAGATTCA AAAGTAGTTC AGCTTTCATA AGTAGAACCA
1201 ATATAAGGAC GTTGTTTTAG CATTTTAAT CATTATTTT AAATAAATGA
1251 TGTAACAGAG GCTTGATTG TGTATGAAA GATTGAGAAA CTAAATTTTC
1301 TGTGTATTTA ATTTTGTGT GCCTTAAAC TTTGTAAAT TCCTGAAGTT
1351 AATTATCATA TTGTACTTTT TGGGACATA CTCATTAGCA GATATGTAGT
1401 GCAGTGATT ACAATAATT GAGAGTAAAA TCAGTGATGT ATAACTAGT
1451 TCATGAGTCT AGGTAAATA TCAATTACCT CTGTTTAAAT TGCTCTGTGA
1501 ATTATTATTG TATGTATTTA AATGTAGTTA AAGCTTTTAA ACATGTTGTT
1551 ACATAGTGTG AATCTACAC AGTGCTACAC AGCTTTTGTG GTCACATAGC
1601 CTTACAGAGT TTATAATGAT GTAGCATCTG CAAAATATAT GCATAGCTTA
1651 TATCCTATTT TTATAGAGCC AGTAATGGTT TTTGTGATGC TGTATTACTT
1701 CTGGGTTTGA GACAATAAAG TCTGTTTAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

```

1  MRLLEMRKQD  WFMVGIVLAI  AGAKLEPSIG  VNGGPLKPEI  TVSYIAVATI
51  FFNSGLSLKT  EELTSALVHL  KLHLFIQIFT  LAFFPATIWL  FLQLLSITPI
101 NEWLLKGLQT  VGCMPPPVSS  AVILTKAVGG  NEAAAIFNSA  FGSFLVSKHS
151 LTLCLLOLL

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phfkd2 4m11, frame 3

PIR:S53951 probable membrane protein YMR034c - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 171, P = 3.2e-12

PIR:A65015 yfeH protein - Escherichia coli (strain K-12), N = 1, Score = 131, P = 4.2e-08

>PIR:S53951 probable membrane protein YMR034c - yeast (*Saccharomyces cerevisiae*)
Length = 434

HSPs:

Score = 171 (25.7 bits), Expect = 3.2e-12, P = 3.2e-12
Identities = 38/144 (26%), Positives = 72/144 (50%)

Query: 5 ERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPKPEITVSYIAVATIFFNSGLSLKTEELT 64
E ++ WE + + + I A+ P+ +GG +K + ++ Y VA IF SGL +K+ L

Shict: 18 EFLKSOWFFICLAILIVIA RFAPNFARDGGLIKGOYSIGYGCVAWIFLOSGLGMKSRSLM 77

Query: 65 SALVHLKLHLFIOIETLAFFPATIWLF---LOLLSITPINEWLLKGLQTVGCMPPPVSSA 121

```

+ + + + + H I + + + + + F + + + I + + W + L GL P V + S

```

Query: 122 VILTKAVGGNEAAAI FNSAFGSFL 145

VI+T	GGN	+	G+ L
------	-----	---	------

Sbjct: 138 VIMTTNAGGNSLLCVCEVFIGNLL 161

Pedant information for DKF2phfkd2_4m11, frame 3

Report for DKFZphfkd2 4m11.3

```
[LENGTH]          159
[MW]               17282.92
[pI]               9.06
[HOMOL]            PR:S53951 probable membrane protein YMR034c - yeast (Saccharomyces cerevisiae)
5e-12
[FUNCAT]           99 unclassified proteins [S. cerevisiae, YMR034c] 2e-13
[PROSITE]          MYRISTYL_2
[PROSITE]          PKC_PHOSPHO_SITE 1
[OWL]              TRANSMEMBRANE_4
```

```
SEQ      MRLLEMRKDFWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSSLTKT
PRD      ccchhhhhhhhhhhhhhhhhhhhhhhccccccccccceeeeecccccccccccchhh
MEM      ....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMM..
```

```
SEQ      EELTSALVHLKLHLFIQIFTLAFFPATIWFLQLLSITPINEWLLKGLQTVGCMPPPVSS
PRD      hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhccccchhhhhhhhheeecccccccc
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
```

```

SEQ      AVILTKAVGGNEAAAI FNSAFGSFLVSKHSLTCLLQLLL
PRD      ceeeecccccchhhhhhccccceeeceeeeeeccc
MEM      mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

```

Prosites for DKFZphfkd2 4m11.3

PS00005	57->60	PKC_PHOSPHO_SITE	PDOC00005
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	129->135	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphfkd2 4m11.3)

PAGE INTENTIONALLY LEFT BLANK

DKFZphut1_17k7

group: uterus derived

DKFZphut1_17k7 encodes a novel 520 amino acid protein with weak similarity to *S. Cerevisiae* Fip1.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to *S.cerevisiae* Fip1

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1914 bp

Poly A stretch at pos. 1897, polyadenylation signal at pos. 1867

```
1 CGGACGCGTG GCGGACGCG TGGGGCTTC CTGGGATTGS AGTCTCGAGC
51 TTTCTTCGTT CTTTCGCCGG CGGGTTCGCG CCCTTCGCG CCCTCGGGGC
101 TGCAGGCGTG GGAAGGGGT TGGAGGGGCG TGTGATCGC CGCGTTAAG
151 TTGCGCTCGG GCGGCCATG TCGGCCGCG AGGTCGAGCG CCTAGTGTG
201 GAGCTGAGCG GCGGACCGG AGGGGATGAG GAGGAAGAGT GGCTCTATGG
251 CGATGAAAAA GAAGTTGAAA GGGCAGAAGA AGAAAAAGCC AGTGCTAATC
301 CTCCATCTGG AATTGAAGAT GAAACTGCTG AAAATGGTGT ACCAAAAACCG
351 AAAGTGACTG AGACCGAAGA TGATAGTGAT AGTGACAGCG ATGATGATGA
401 AGATGATGTT CATGTCACTA TAGGAGACAT TAAACCGGA GCACCAACAGT
451 ATGGGAGTTA TGGTACAGCA CCTGTAAATC TTAACATCAA GACAGGGGGA
501 AGAGTTTATG GAATACAGG GACAAAAAGT AAAGGAGTAG ACCTTGATGC
551 ACCTGGGAGC ATTAATGGAG TTCCACTCTT AGAGGTAGAT TTGGATTCTT
601 TTGAAGATAA ACCATGGCGT AAACCTGGTG CTGATCTTTC TGATTATTTT
651 AATTATGGGT TTAATGAAGA TACCTGGAAA GCTTACTGTG AAAACAAAAA
701 GAGGATACGA ATGGGACTTG AAGTTATACC AGTAACCTCT ACTACAAAAA
751 AAATTACGGT ACAGCAGGGA AGAACTGGAA ACTCAGAGAA AGAACTGCC
801 CTTCATCTA CAAAGCTGA GTTACTTCT CCTCTCTCT TGTTCAGGAC
851 TGGGCTTCCA CCGAGCAGGA GATTACCTGG GGCATTGAT GTTATCGGTC
901 AGACTATAAC TATCAGCCGA GTAGAAGGCA GGCAGCGGG AAATGAGAAC
951 AGCAACATAC AGGTCCTTTC TGAAGATCT GCTACTGAAG TAGACAACAA
1001 TTTAGCAAA CCACCTCCGT TTTTCCCTCC AGGAGCTCCT CCCACTCACC
1051 TTCCACCTCC TCCATTCTT CCACCTCCTC CGACTGTCAG CACTGCTCCA
1101 CCTCTGATTC CACCACCGGG TTTTCTCTCT CCACCGGCG CTCACCTCTC
1151 ATCTCTTATA CCAACAATAG AAAGTGGACA TTCCTCTGGT TATGATAGTC
1201 GTTCTGCACG TGCATTTCCTA TATGGCAATG TTGCTTTCC CCATCTTCTT
1251 GGTTCGCTC CTTCTGGGCC TAGTCTTGTG GACACCGCA AGCAGTGGGA
1301 CTATTATGCC AGAAGAGAGA AAGACCGAGA TAGAGAGAGA GACAGAGACA
1351 GAGAGCGGGA CCGTGATCGG GACAGAGAAA GAGAACGCAC CAGAGAGAGA
1401 GAGAGGGGAG GTGATCACAG TCCTACACCA AGTGTTTTCA ACAGCGATGA
1451 AGAACGATAC AGATACAGGG AATATGCAGA AAGAGGTTAT GAGCGTCACA
1501 GAGCAGTCG AGAAAAAGAA GAACGACATA GAGAAAGACG ACACAGGGAG
1551 AAAGAGGAAA CCAGACATAA GTCTTCTCGA AGTAATAGTA GACGTCGCCA
1601 TGAAGTGAA GAAGGAGATA GTCACAGGAG ACACAAACAC AAAAAATCTA
1651 AAAGAAGCAA AGAAGGAAAA GAAGCGGGCA GTGAGCCTGC CCCTGAACAG
1701 GAGAGCACCG AAGTACACC TGCAGAATAG GCATGGTTT GGCCTTTTGT
1751 GTATATTAGT ACCAGAAGTA GATACTATAA ATCTTGTTAT TTTTCTGGAT
1801 AATGTTTAAG AAATTACCTT TAAATCTGT TCTGTTGTT AGTATGAAAA
1851 GTTAACCTTT TTTCCAAAT AAAAGAGTGA ATTTTTCATG TTAAGTTAAA
1901 AAAAAAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 168 bp to 1727 bp; peptide length: 520
Category: similarity to known protein

```
1 MSAGEVERLV SELSGGTGGD EEEEWLYGDE NEVERPEEEN ASANPPSGIE
51 DETAENGVPK PKVTETEDDS DSDSDDEDD VHTIGDIKT GAPOYGSYGT
101 APVNLNIKTG GRVYGTGTGK VKGVDLDAPG SINGVPLLEV DLDSFEDKPW
151 RKPGLDSY FNYGFNEDTW KAYCEKQKRI RMGLEVIPVT STTNKITVQQ
201 GRTGNSEKET ALPSTKAEFT SPPSLFKTGL PPSRRLPGAI DVIGQITIS
251 RVEGRRRANE NSNIQVLSEK SATEVDNNFS KPPFFFPFGA PPTHLPPIPF
301 LPPPTVSTA PPLIPPPGFP PPPGAPPPSL IPTIESGHSS GYDSRSARAF
351 PYGNVAFPHL PGSAPSWPSL VDTSKWDYY ARREKDRDR RDRDRERDR
401 RDRERERTRE RERERDHSPT PSVFNSDEER YRYREYAERG YERHRASREK
451 EERHRERRHR EKEETRHKSS RSNSRRRHES EEGDSHRRHK HKKSKRSKEG
501 KEAGSEPAPE QESTEATPAE
```

BLASTP hits

Entry AF016427.4 from database TREMBL:
gene: "F32D1.9"; *Caenorhabditis elegans* cosmid F32D1.
Score = 392, P = 1.8e-36, identities = 156/519, positives = 212/519

Entry S62454 from database PIR:
hypothetical protein SPAC22G7.10 - fission yeast (*Schizosaccharomyces pombe*)
Score = 246, P = 2.0e-22, identities = 62/163, positives = 91/163

Entry A56545 from database PIR:
PII1 protein - yeast (*Saccharomyces cerevisiae*)
Score = 186, P = 2.9e-16, identities = 56/206, positives = 92/206

Alert BLASTP hits for DKFZphut1_17k7, frame 3

TREMBLNEW:AF109907.1 product: "S164"; *Homo sapiens* S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds., N = 2, Score = 236, P = 1.5e-16

>TREMBLNEW:AF109907.1 product: "S164"; *Homo sapiens* S164 gene, partial cds; PS1 and hypothetical protein genes, complete cds; and S171 gene, partial cds.

Length = 735

HSPs:

Score = 236 (35.4 bits), Expect = 1.5e-16, Sum P(2) = 1.5e-16
Identities = 51/120 (42%), Positives = 76/120 (63%)

Query: 383 REKDRDRERDRDRDRDRERERTREERERERDHSPTPSVFNSDEERYRYREYA---ER 439
REK+++RER+R+R+RDRDR +ER+R R+RER+RD S + +++R R RE + ER
Sbjct: 227 REKEKERERERERDRDRDRDKERDRDRDRDRDRDRERSS-DRNKDRSRSEKSRDRER 285

Query: 440 GYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKSKRSK 498
ER R + ER RER R RE+E R + + +R E +E D++ R K ++ R K
Sbjct: 286 EREREREREREREREREREREREREREREREKDKKADREDEEDAYERRKLEKLEK 345

Query: 499 E 499
E
Sbjct: 346 E 346

Score = 214 (32.1 bits), Expect = 4.4e-14, Sum P(2) = 4.4e-14
Identities = 50/133 (37%), Positives = 75/133 (56%)

Query: 383 REKDRDR-ERDRDRERDRDRDRERERTREERERERDHSPTPSVFNS-DEERYRYREYAERG 440
RE++R+R ER+R+RER+R+R++E+ER RERER+RD T D ER R R+ ER
Sbjct: 208 REREREREREREREREREREKEKERERERERDRDRDRDKERDRDRDRDRDRDR-RERS 266

Query: 441 YERHRASREKEERHRERHRREKEETRHKSSRSNSRRRHESEEGDSHRRHKHKSKRSKEG 500
+R++ E+R+R RE+E R+ R R R E + R + ++ K K
Sbjct: 267 SDRNKDRSRSEKSRDRERERERERERE-REREREREREREREREREREKDKKRD 324

Query: 501 KEAGSEPAPEQESTE 515
+E E A E+ E
Sbjct: 325 REEDEEDAYERRKLE 339

44i

Query: 492 -----KKSRSKEGKEAGSEPAPEQESTE 515
+K R +E + E ++E E
Sbjct: 405 YRGSALQKRLRDREKEMEADERDRKREKEE 434

Score = 162 (24.3 bits), Expect = 2.4e-08, Sum P(2) = 2.4e-08
Identities = 45/141 (31%), Positives = 74/141 (52%)

Query: 372 DTSKQWDYARREKDRDRDRDRDRDRDRDRERTRERERERHDSPTPSVFNSEERY 431
+ SK D + + E+++ ++ +E +++R RERER RERERER + ER
Sbjct: 172 EISKFRDTHKKLEEEKGKKEKERQEIEKER-RERERERERERERERERERERERERERE 228

Query: 432 RYREYAERGVERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHK 490
+ +E ER ER R +ER R+R R R+++ R +SS N R E+ R +
Sbjct: 229 KEKE-RERERERDRDRDKERDRDRDRDRDRDRDRDRSSDRNKDRSRSEKSRDRERER 287

Query: 491 HKKSRSKEGKEAGSEPAPEQE 512
++ +R +E +E E E+E
Sbjct: 288 ERERERERE-RERERERERERE 308

Score = 137 (20.6 bits), Expect = 1.2e-05, Sum P(2) = 1.2e-05
Identities = 48/152 (31%), Positives = 68/152 (44%)

Query: 364 APSWPSLVDTSKQWDYARREKDRDR-ERDRDRDRDRDRDRDRERTRERERERHDSPTPS 422
AP P + T + + E RD R+ + RD + E E+ + +E+ER
Sbjct: 143 APLIPYPLITKEDINAIMEEDKRLISREISKFRDTHKKLEEEKGK-KEKERQEIEKER 201

Query: 423 VFNSDEERYRYREYAERGVERHRA-SREKE-ERHRER-RHREKEETRHKS-SRSNSRRRH 478
+ ER R RE ER ER R REKE ER RER R R+++ T+ + R R R
Sbjct: 202 R-EREREREREREREREREREREREREKEKERERERERDRDRDRDKERDRDRDRDRD 260

Query: 479 ESEEGDSHRRHKKKSRSKEGKEAGSEPAPEQE 512
E S R +S+ +E E E+E
Sbjct: 261 RDRRSSDRNKDRSRSEKSRDRERERERERE 294

Score = 126 (18.9 bits), Expect = 1.8e-04, Sum P(2) = 1.8e-04
Identities = 41/149 (27%), Positives = 66/149 (44%)

Query: 375 KQWDYARREKDRDRDRDRDRDRDRDRERTRERERERHDSPT---PSVFNSD--EE 429
K W+ R+K R+ E++ +RE +R R+ +E R +E D+ P + + +
Sbjct: 354 KNWEI-RERKKTREYEKEAEEREEERRREMAKEAKRLKEFLEDYDDDRDDPKYRGSALQK 412

Query: 430 RYRYREYAERGVERHRASREKEERHRER-----HREKEETRHKSSRSNSRRRHES--E 481
R R RE ER R REKEE R+ H + + + + RRR +
Sbjct: 413 RLDRREKEMEADERDR-KREKEELEIRQLLAEGHPDPAELQRMQEAEARRRQPOIKQ 471

Query: 482 EGDHRRHKKKSRSKEGKEAGSEPAPEQE 512
E +S + K+ K K + E EEO+
Sbjct: 472 EPESEEEEEKQKEKEEKREPEMEEEEEPEQK 502

Score = 124 (18.6 bits), Expect = 3.0e-04, Sum P(2) = 3.0e-04
Identities = 41/141 (29%), Positives = 65/141 (46%)

Query: 380 YARREKORD-RERDRDRDRDRDRDRDRERTRERERERHDSPTPSVFNSEERYRYREYAE 438
Y R K+ + RER + RE +++ +RE ER RE +E + + D++R + Y
Sbjct: 349 YQERLKNWEIREKKTREYEKEAEEREEERRREMAKEAKRLKE-FLEDYDDDRDDPKYRGSALQK 407

Query: 439 RGYERHRASREKEERHRER-RHREKEETRHKSSRSNSRRRHESEEGDSHRRHKKKSRS 497
++ REKE ER R REKEE R + H + + R + + +R
Sbjct: 408 SALQKRLDRREKEMEADERDRKREKEELEIRQLLAEG-HPDPAELQRMQEAEARRRQ 466

Query: 498 KEGKEAGSEPAPEQESTATPAE 520
+ K+ EP E+E E E
Sbjct: 467 PQIKQ---EPESEEEEEKQKE 486

Score = 121 (18.2 bits), Expect = 6.2e-04, Sum P(2) = 6.2e-04
Identities = 43/149 (28%), Positives = 67/149 (44%)

Query: 364 APSWPSLVDTSKQWDYARREKDRDR-ERDRDRDRDRDRDRDRERTRERERERHDSPTPS 422
AP P + T + + E RD R+ + RD + E E+ + +E+ER
Sbjct: 143 APLIPYPLITKEDINAIMEEDKRLISREISKFRDTHKKLEEEKGK-KEKERQEIEKE- 200

Query: 423 VFNSDEERYRYREYAERGVERHRASREKEERHRERHREKEETRHKSSRSNSRRRHESEE 482
+ ER R RE ER R RE+E + R RE+E R + R+ R R E
Sbjct: 201 --RREREREREREREREREREREREREKEKERERERERDRDRDR-RTKERDRDRDRE 256

Query: 483 GDSHRRHKKKSRSKEGKEAGSEPAPEQE 512
D R + + S R+K+ + E + +E
Sbjct: 257 RDRDR-DRRSSDRNKD-RSRSEKSRDRE 284

Score = 105 (15.8 bits), Expect = 3.1e-02, Sum P(2) = 3.1e-02

Pedant information for DKFZphutet 17k7, frame 3

Report for DKFZphutel 17k7.3

443


```
PRD      hhhheeecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      DVIGQTITISRVEGRRRANENSNIQVLSERSATEVDNNFSKPPPPFFPGAPPTHLPPIPPFF
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      LPPPTVSTAPPLIPPGFPPPGAPPPSLIPTIESGHSSGYDSRSARAFYPYGNVAFPHL
SEG      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PGSAPSWPSLVDTSKQWDYYARREKDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      PSVFNSEERYRYREYAERGERHRASREKEERHRERRRREKEETRHKSSRSNSRRRHES
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      EEGDSHRRHKHKKSKRSKEGSEPAPEQESTATPAE
SEG      xx..XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphut1_17k7.3

PS00001	40->44	ASN_GLYCOSYLATION	PDOC00001
PS00001	278->282	ASN_GLYCOSYLATION	PDOC00001
PS00005	169->172	PKC_PHOSPHO_SITE	PDOC00005
PS00005	193->196	PKC_PHOSPHO_SITE	PDOC00005
PS00005	206->209	PKC_PHOSPHO_SITE	PDOC00005
PS00005	214->217	PKC_PHOSPHO_SITE	PDOC00005
PS00005	233->236	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	373->376	PKC_PHOSPHO_SITE	PDOC00005
PS00005	469->472	PKC_PHOSPHO_SITE	PDOC00005
PS00005	474->477	PKC_PHOSPHO_SITE	PDOC00005
PS00005	485->488	PKC_PHOSPHO_SITE	PDOC00005
PS00005	494->497	PKC_PHOSPHO_SITE	PDOC00005
PS00006	2->6	CK2_PHOSPHO_SITE	PDOC00006
PS00006	17->21	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	70->74	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	215->219	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	273->277	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	426->430	CK2_PHOSPHO_SITE	PDOC00006
PS00007	434->442	TYR_PHOSPHO_SITE	PDOC00007
PS00007	152->161	TYR_PHOSPHO_SITE	PDOC00007
PS00008	15->21	MYRISTYL	PDOC00008
PS00008	96->102	MYRISTYL	PDOC00008
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	154->160	MYRISTYL	PDOC00008
PS00008	229->235	MYRISTYL	PDOC00008
PS00008	244->250	MYRISTYL	PDOC00008
PS00008	289->295	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00009	253->257	AMIDATION	PDOC00009

(No Pfam data available for DKFZphut1_17k7.3)

DKF2phut1.18c12

group: uterus derived

DKF2phut1.18c12 encodes a novel 378 amino acid protein nearly identical to human WUGSC:H_DJ0872F07.1 protein.

The novel protein has an additional N-terminal domain, which is not present in WUGSC:H_DJ0872F07.1.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

nearly identical to human WUGSC:H_DJ0872F07.1 protein

on genomic level encoded by AC004537, 10 exons the predicted protein sequence AC004537.1 is only partially o.k. first exon wasn't predicted there are additional exons predicted (BLASTX/EST-BLAST shows that the cDNA is only partly spliced) intron -1216-3540//~3577-5059

Sequenced by AGOWA

Locus: map="7q31"

Insert length: 6005 bp

Poly A stretch at pos. 5980, polyadenylation signal at pos. 5968

```

1  AGCGGGTGCT GCTAGCGGAG GCGCCATATT GGAGGGGACA AAACCTCCGGC
51  GACAGCGAGT GACACAAATA AACCCCTGGA CCCCTTGTT CCCTCAGCTC
101 TAAGGGCCGC GATGTTGTAC CTAGAAGACT ATCTGGAAAT GATTGAGCAG
151 CTTCTATGCG ATCTGCGGGA CCGCTTCACG GAAATGCGCG AGATGGACCT
201 GCAGGTGCAG AATGCAATGG ATCAACTAGA ACAAAGAGTC AGTGAATTCT
251 TTATGAATGC AAAGAAAAAT AAACCTGAGT GGAGGGGAGA GCAATGGGCA
301 TCCATCAAAA AAGACTACTA TAAAGCTTTG GAAGATGCAG ATGAGAAAGT
351 TCAGTTGGCA AACAGATAT ATGACTTGGT AGATCGACAC TTGAGAAAGC
401 TGGATCAGGA ACTGGCTAAG TTTAAATGG AGCTGGAAAG TGATAATGCT
451 GGAATTACAG AAATATTAGA GAGGCGATCT TTGGAATTAG ACACCTCTTC
501 ACAGCCAGTG AACAAATCACC ATGCTCATTC ACATACTCCA GTGGAAAAAA
551 GGAAATATAA TCCAACCTCT CACCATACGA CAACAGATCA TATTCCTGAA
601 AAGAAATTTA AATCTGAAGC TCTTCTATCC ACCCTTACGT CAGATGCCTC
651 TAAGGAAAAA ACACCTAGGT GTCGAAATAA TAATCCACA GCCTCTCTTA
701 ACAATGCCCT CAATGTGAAT TCCTCCCAAC CTCTGGGATC CTATAACATT
751 GGCTCGTTAT CTTGAGAAC TGGTGCAGGG GCAATTACCA TGGCAGCTGC
801 TCAAGCAGTT CAGGCTACAG CTCAGATGAA GGAGGGGACGA AGAACATCAA
851 GTTTAAAGC CAGTTATGAA GCATTTAAGA ATATGACTT TCAGTTGGGA
901 AAAGAAATTT CAATGGCCAG GGAACAGTT GGCTATTCAT CATCTTCGGC
951 ACTTATGACA ACATTAACAC AGAATGCCAG TTCATCAGCA GCCGACTCAC
1001 GGAGTGGTCG AAAGAGCAAA AACAAACAAC AGCTTCAAG CCAGCAGTCA
1051 TCATCTCTCT CCTCCTCTTC TTCTTTATCA TCGTGTCTCT CATCATCAAC
1101 TGTGTACAAA GAAATCTCTC AACAAACAAC TGTAAGTGCCA GAATCTGATT
1151 CAAATAGTCA GGTGATGTC ACTTACGACC CAAATGAACC TCGATCTGTC
1201 ATTTGTAATC AGGTAAAAGT CTGTTATATC TATAAAAGTA TAATCTGAAT
1251 AAACCTAGAA GAAGAGAAGT ATTTCAATTT TAAGCACTTT TTTAACTCA
1301 CTTAAATATC CTTTGCTTTA TTTGTATACT TTTCTCCCCC TTCTTACAAA
1351 AGTGACATTT GCTGTAAATA CTGAGTATAA AGAAAAATGT TACCCATAAT
1401 CCTAGCCCTC AGATACAACC TGTAACATAA CATTTTTGGT ATACCACTAC
1451 CATATACCTC ATGTGCACAT TGGCTGCCTT AATAAAATAC AACAGACTGG
1501 GTAGCTTAAA CAACAGAAAA TAATTTCTCT ACAGGTATGA AGGCTGGGAA
1551 GTCCAAGATC AAGGTGTCCA CTGACTCAGT TCTGGAGGAG GGCTCCCTTC
1601 CTAGATGGAG ACTGCTGCTT TCTCACCAGG TCCTCAGATG ATAGAGGGAG
1651 AAAGAGTGTG CTCTGGTGTG TTTTCTTATA AGGGCACCAG CCTTGTCAGA
1701 GTAGGACCCC ACTCTATGAC CTCATTTAAC CTTTACCACC TCCTCACAGG
1751 CCCTGTTTCC AATTATAGTC ACGTTGGGGG TTAGGGCTTC AACATATGAT
1801 TTTGAGACAT AAGCTTGCA TTTCAATCAC GTGTCTATGC AGATTTCAC
1851 ATGCATGTGT GTATAAGTTT GTCAGTAGGA ACCACAGTGT ATACTTTCTT
1901 GTTACTGGCT TTTTCTCTA AATCAGGTAT ACCGAACATG ATTTTCTTT
1951 AAGATCATAT TTTTAATTTT CACATAGTTA TCTCTTATGC CATCCAGTGT
2001 AGTTTCTCTA ACCAATACCT AGCTATAGAT TATATTAGTG GTTTTAATTT
2051 GTTTGAAATT AGGGAATAA TTACGATAGG CATTTTTAA ATGTAATCCA
2101 TTTTATACAT CTAATTTCTT GGATAATCTT TTAGAAATAA AATTAGGCTG
2151 TAAATATTGT ACAGACACCA AAATATATTT TCTAGAAATT TATTACCAAA
2201 AATTAAATAA CATACCGGTT TACTAAACCC TGTCCAACAC TGGATATTAT
2251 TTTCTTTAA AACTAAGTA CCAATTGGT AGTTTATAT TATGATTGTT
2301 TTAATACAC TAGTATTAT GAAGTTGGAC ATTTTGTGAC CATTTTGTT
2351 TTTTACATTA TGAATCGACT CCTAATGGT TCGGCTGATT TTTCTATTGT

```

```
2401 TTTTGTATG TACTCTAAAT ATTTGCTTGA TTTAGTTTTT TAAAAATAAT
2451 TCTAAAAATTT TAATTTTATG TAGTTATGAC TGTAAATTTT TTTTATGAA
2501 GCAGGCCATG GATTATATAC TTAGAAGGCG TTTCTCTTTG GCTCTCTTT
2551 CTACAAAAAA TTGCTCTGTA TAATATTTTC TCCTAGTTTT TATATGGTTT
2601 TGTCTAGTTC TTTGCTGCTC TCAGTTTCTT CACATTTAAG ACTTAGTCTA
2651 TCAGCAGATT ATTTGTGCTA ACAGATGAGT TTGCCAGTCT GATTTTTAAA
2701 AATTTTAACA ATTTGTTAGC TGTCCACTA TCACCCGATA AACATTTTTC
2751 AGTACAAATG ATAGAAAAGC ATATCTCTGA TCCTGACAAC AAAAGTAGAT
2801 TACTTGCAAA AGAACAAAAT CAGACTGAAC CTAGAGTTTT CCTCTGTAAC
2851 ACTAAAAAAC TAGAAGGTGA TGGAAATATG CTGTAGAGCT TTCAGGGAAA
2901 AATTAAGAGC CCCCAAAAAC TTGATATTCA GAGAAGTTAT TTCTCTGCAT
2951 AGGACCATTG AATATATATT TCACATCATG AGAAGATCAG AAGATATGCC
3001 ATCTAGTTAA TCCTGCTCGA AAAATTATTC AATCCACTGA GAACTTCAGT
3051 GAACTCAAGA ATTAGCAAGT TATGCCCTAA AGTGCTGGTG ATGAAGAGCA
3101 AAGAAAAAAT GAGAAAGGAC ATAAAAAGA TAAGTTTAGA AGTTTCAAGG
3151 AAGGAGACTA TTAATTGCAA AATATATAT GACCTAATGT GACCCAAGAA
3201 GTAAAAACTT TCAGTAAGTA AATATCAAG AAAGGAACCT AAAATTTTTA
3251 CAATAAGAAC TACCAGAGAA GATGACTCCT TCATCCGGGT GATTATATG
3301 TCAAGTTCTT CCAGACTTCT GAAGGGCAGA TAATTCCTGT GCATTTCTTC
3351 CCACCCCTGC CCCACCTCTG CCAAAAGAGT ATTTGAGGAA AAAATTTATTA
3401 TACCTTGATT CTCAATGTAA TTGTATATTC AGTGATTTTC CCTTTATTTT
3451 CCAGCAGTAT CATACATAAA CAGTTAATTG GTATCTAGGT GTTTGTACAA
3501 TAGTCATAAT AAAGACATTT AATTTTTTTT AACTAGGTAT CTTATGGTGA
3551 GATGGTGGGA TGTGATAACC AAGATGTAAG TATTACATT TTCTATTAG
3601 GAATGAAAAA AATCAGAGT TGTATTACT TGAATATTG TCTATTGTC
3651 TGTATGGTTT GGTCTAAGAA AACAGGTTG CAGGTATATT AGTTATGTTA
3701 TGCTAATGCT AGAATATTCC TCTCAAAAT AGGGTAGTGT CCCTTAATGT
3751 GTTCCCTATT TTAATTTTTA AAGCTAATT TATGGTTTTA TGTGAGATT
3801 GTCTCAGAAG TGTATGTTG TATGAAATTT ATAAATACCC TCCTTTCCCT
3851 TTAATAAAAA ATACTGTGTT TACTAGAATC CAGTTCATT ATCAGATTGA
3901 AGAAATGGAA TTTTAAACA ATTCATTCT TCAGGCTGCA CCGTCTTAAA
3951 GTGAAGGGTG GGAATAATGA GGATCTAATG TGAGATTATC TTCTCTCAT
4001 GAGTATAATA TTTTTCCCTG TACTCTGAG GTGTCAGCTG ATAAGAGCCA
4051 CCCCTGATCT AAAAAAGTAA GGAATTTGA AAGGAAGGAA TTCTTGGTTT
4101 TTAGGAGACT TAATTTTAGT TAGAGATACG TTTTATTATC AATCTGAGA
4151 ATATTGTTGT CTAGTAATTT TGACTTCCCTC CTTATTTAGT AGTGACAGGA
4201 TCCTAAGATT AACAAAGATT TTAATTTTGT AAACAATCT CAAGATTCAG
4251 GGAGCTGSGT AGGTGCAATT AAATGCTGAT TTTCTTAAGA CCTGATAGG
4301 TTACAGCAAC ATGTCACCT AGATTGGGAG AGAGCTCTCT TCTGTTTCCC
4351 TCTCTAGAA CCCTTGAGG CTGTTTCTGG TTGTTGCAAA AACATATTG
4401 CCCACCAATT TCAAGAACAT CACTCTAACC TCTTCTGGG CAGTATGTA
4451 AATGATGAA TGAGATTCT ATGACTACCA GCATCATGCT TCTCTGATTC
4501 TTCTTATTC CAGTTCTGCT CTCTGAGTG CTAAGACTTT CATGAAGAG
4551 TTTTCTGCTT AATATGTTT AAAGAGGAAT AATTTTCTC TACATTTCAA
4601 GGAATAGAAA CACCACGTA GGAATGAGC GGCATAGAC ATAAATTAAT
4651 GTCTTTAATT ACAATCAGCT TATTCTACTT TATGAGACAG CAAATAGGC
4701 TGACTATTAA ATAAATCTT AAGTTATATT TACCTCTAC ATAGAAGATT
4751 CATCCCACTT CTTTTGCCCT TTGAAAGCTG AAAACTAGTG AATTTTCATT
4801 CATTAGGATG AGGGGACTAG ATTACATGGA CCTCAGGATT CTTGAAGATG
4851 CATAATTTTT CTGTGCCTTC ATTTCTCAT TCCTGAAGCT TATCATTTAG
4901 TCTAATGAT GTCTAATAA TCTAGATCTA AAAATCTCTA GTGCACACAT
4951 CTAATTATTG TTAATTTAAA TGGATTATTC AGTCTCTCTA GCATATTTTA
5001 ATATACTCTC TTGTCTTCAG AAGTACTGAA AACTTGTTTT TTGCAATTTT
5051 GCTTTCTAGT GCCCTATAGA ATGGTCCAT TATGGCTGCG TTGCAATTGAC
5101 AGAGGCCACA AAAGGCAAAAT GGTACTGTC CACAGTGCAT GCTGCAATGA
5151 AGAGAGAGGG CAGCAGACAC AAATAAAGGT GGTCTTTTGT TTTGATGAAG
5201 AATAAAGCTT CAGCTGAAGA TTTTATATAG GACTTTAAAA AGAAGAGAAG
5251 AGAAAGAAGA AACAAATGCAT TTCCAGGCAA CCACTTAAAG GATTTACATA
5301 GACAATCCTA TAAGATCTTG AACTTGAATT TTAGGGTGT TATTTAATA
5351 ATGTAAGTAA ATTATTTATG CACTCCTGGT GTGCTATGAA TATTATTCCA
5401 GTTAGCCTTG GATTATTCCA GTGGCCAACA TATGCAGACA TTTGTACTCC
5451 TCAACCATTT TCTCAAAGTA ATGGGCATTC TATGATTAG ACTTCAAGGA
5501 ATTCCAATGA TGAAGATTTT AAGGAAAGTA TTTTATATTC AACAGGTATA
5551 TTCTGCTGCA TGTACTGTAC TCCAGAGCTG TTATGTAACA CTGTATATAA
5601 ATGGTTGCAA AAAAAAATAA AAGTCAGTGC TTCTAAAAAG AATTTAAGAT
5651 AATGGTTTTT AAAATGCCCT TATAATAAGC TTTGTTTCT TGTGAAACTA
5701 ATTCAGCAGG CTGAAGGAAA TGGTTATGAT GATAATGTGG GCTGGTATCC
5751 TCTAGAGTAC CTGGGTACAT AAACAGAAAC TCCTGTAGGT AAAAAGTAAT
5801 TTGTGCCATT AGTCTTCTTA TGTTCCTGCA TCCAGATAGA GTGCAGTTCA
5851 TGAGGGAGGG GCGGGGGGAC TGAAGGGGAA AGGGCGTTAA AGTGATACAT
5901 TTTTATACCA AATGTGTTTA TTTTGTGTG CAAGTAATCC TTAATAATGC
5951 AATTGTATTA GGTGTTAAAA TAAAGTTTTT AAAAAATTAA AAAAAAATAA
6001 AAAAA
```

BLAST Results

Entry HSG20547 from database EMBL:
HSG20547| human STS A005W09.
Length = 154

Minus Strand HSPs:
 Score = 770 (115.5 bits), Expect = 2.9e-26, P = 2.9e-26
 Identities = 154/154 (100%)

Medline entries

98101645:
 The candidate tumour suppressor p33ING1 cooperates with p53 in cell growth control.

Peptide information for frame 1

ORF from 112 bp to 1245 bp; peptide length: 378
 Category: similarity to known protein

```

1 MLYLEDYLEM IEQLPMDLRD RFTEMREMDL QVQNAMDQLE QRVSEFFMNA
51 KKNKPEWREE QMASIKKDYI KALEDADEKV QLANQIYDLV DRHLRKLQDE
101 LAKFKMELEA DNAGITEILE RRSLELDTPS QPVNNHHAHS HTPVEKRRYN
151 PSHHTTTTDM IPEKKFKSEA LLSTLTSDAS KENTLGCRRN NSTASSNNAY
201 NVNSSQPLGS YNIGSLSSGT GAGAITMAAA QAVQATAQMK EGRRTSSLKA
251 SYEAFKNNDF QLGKEFSMAR ETVGYSSSSA LMTTLTQNAS SSAADSRSGR
301 KSKNNNNKSSS QSSSSSSSSS SLSSCSSST VVQEISQQT VVPESDSNSQ
351 VDWTYDPNEP RYCICNQVKV CYIYKSI

```

BLASTP hits

Entry AF044076.1 from database TREMBL:
 "ING1"; product: "candidate tumor suppressor p33ING1"; Homo sapiens candidate tumor suppressor p33ING1 (ING1) mRNA, complete cds. Homo sapiens (human)
 Length = 279
 Score = 162 (57.0 bits), Expect = 1.1e-09, P = 1.1e-09
 Identities = 48/183 (26%), Positives = 92/183 (50%)

Entry AC004537.1 from database TREMBL:
 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07 from 7q31, complete sequence.
 Score = 1814, P = 3.7e-187, identities = 358/358, positives = 358/358

Entry CEY51H1A.1 from database TREMBL:
 gene: "Y51H1A.4"; Caenorhabditis elegans cosmid Y51H1A
 Score = 213, P = 3.7e-15, identities = 37/123, positives = 82/123

Alert BLASTP hits for DKFZphut1_18c12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18c12, frame 1

Report for DKFZphut1_18c12.1

```

[LENGTH]      378
[MW]           42275.72
[pI]           5.72
[HOMOL]        TREMBL:AC004537.1 gene: "WUGSC:H_DJ0872F07.1"; Homo sapiens PAC clone DJ0872F07
from 7q31, complete sequence. 1e-157
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YHR090c] 8e-05
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YNL097c] 2e-04
[PROSITE]      MYRISTYL 3
[PROSITE]      AMIDATION 2
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 4
[PROSITE]      PROKAR_LIPOPROTEIN 1
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 3
[PROSITE]      ASN_GLYCOSYLATION 5
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY 20.63 %

```

[illegible]

Prosites for DKFZphute1_18c12.1

PS000001	190->194	ASN GLYCOSYLATION	PD0C00001
PS000001	191->195	ASN GLYCOSYLATION	PD0C00001
PS000001	203->207	ASN GLYCOSYLATION	PD0C00001
PS000001	288->292	ASN GLYCOSYLATION	PD0C00001
PS000001	306->310	ASN GLYCOSYLATION	PD0C00001
PS000002	218->222	GLYCOSAMINOGLYCAN	PD0C00002
PS000004	243->247	CAMP_PHOSPHO_SITE	PD0C00004
PS000005	64->67	KFC_PHOSPHO_SITE	PD0C00005
PS000005	247->250	KFC_PHOSPHO_SITE	PD0C00005
PS000005	298->301	KFC_PHOSPHO_SITE	PD0C00005
PS000006	142->146	CK2_PHOSPHO_SITE	PD0C00006
PS000006	156->160	CK2_PHOSPHO_SITE	PD0C00006
PS000006	292->296	CK2_PHOSPHO_SITE	PD0C00006
PS000006	349->353	CK2_PHOSPHO_SITE	PD0C00006
PS000008	186->192	MYRISTYL	PD0C00008
PS000008	214->220	MYRISTYL	PD0C00008
PS000008	249->225	MYRISTYL	PD0C00008
PS000009	211->245	AMIDATION	PD0C00009
PS000009	298->302	AMIDATION	PD0C00009
PS000013	315->326	PROKAR_LIPOPROTEIN	PD0C00013

(No Pfam data available for DKFZphut1_18c12.1)

DKFZphutel_18i19

group: transcription factors

DKFZphutel_18i19 encodes a novel 759 amino acid protein with similarity to the SREBP-2 mutant sterol regulatory element binding protein-2 of *Cricetulus griseus*.

The SREBP-2 protein is embedded in the membranes of the nucleus and endoplasmic reticulum. In cholesterol-depleted cells the proteins are cleaved to release soluble NH2-terminal fragments that enter the nucleus and activate genes encoding the low density lipoprotein receptor and enzymes of cholesterol synthesis. The new protein is a putative transcription factor capable of protein-protein interaction via a lim domain and additionally shows similarity to the common sunflower transcription factor SF3.

The new protein can find application in modulating/blocking the expression of genes involved in lipid metabolism.

similarity to transcription factor SF3

complete cDNA, complete cds, EST hits
strong similarity to mutated SREBP-2 of hamster,
similarity is not to SREP-2 part of protein but to the unknown part of
the fusion protein

Sequenced by AGOWA

Locus: /map=12

Insert length: 3664 bp

Poly A stretch at pos. 3647, polyadenylation signal at pos. 3636

```
1 GCGCTAGGTA GAGCGCCGGG ACCTGTGACA GGGCTGGTAG CAGCGCAGAG
51 GAAAGGCGGC TTTAGCCAG CTATTTCAGT GTCTGTAGAC AAGATGSAAT
101 CATCTCCATT TAATAGACGG CAATGGACCT CACTATCATT GAGGCTAACA
151 GCCAAGAAGC TTTCTCTTCT CAACAAGAAC AAGTCATCGG CTATTGTGGA
201 AATATTCTCC AAGTACCAGA AAGCAGCTGA AGAACAACAA ATGGAGAAGA
251 AGAGAAGTAA CACCGAAAT CTCTCCGAGC ACTTTAGAAA GGGGACCTGT
301 ACTGTGTTAA AGAAGAAGTG GGAGAACCCA GGGCTGGGAG CAGAGCTCA
351 CACAGACTCT CTACGGAACA GCAGCAGCTG GATTAGGCAC AGAGCAGACC
401 ATCTCTCTGC TGAAGTGACA AGCCACGCTG CTTCTGGAGC CAAAGCTGAC
451 CAAGAAGAAC AAATCCACCC CAGATCTAGA CTCAGGTCAC CTCCTGAAGC
501 CCTCGTTTCA GGTGATATC CCCACATCAA GGACGGTGAG GATCTTAAAG
551 ACCACTCAAC AGAAAGTAAA AAAATGGAAA ATTGTCTAGG AGAATCCAGG
601 CATGAAGTAG AAAAATCAGA AATCAGTGAA AACACAGATG CTTGGGGCAA
651 AATAGAGAAA TATAATGTTC CGCTGAACAG GCTTAAGATG ATGTTTGAGA
701 AAGGTGAACC AACTCAAACT AAGATTCTCC GGGCCCAAAG CCGAAGTGCA
751 AGTGAAGGGA AGATCTCTGA AAACAGCTAT TCTCTAGATG ACCTGGAAAT
801 AGGCCCCAGG CAGTTGTCTA CTTCTACTAT TGAATCGGAG AAAAATGAGA
851 GTAGACGAAA TCTGGAACCT CCACGCCCTC CAGAAACCTC TATAAAGGAT
901 CGAATGGCCA AGTACCAGGC AGCTGTGTCC AAACAAAGCA GCTCAACCAA
951 CTATACAAAT GAGCTGAAAG CCAGTGGTGG CGAAATCAAA ATTCATAAAA
1001 TGGAGCAAAA GGAGAATGTG CCCCAGGTC CTGAGGTCCT CATCACCCAT
1051 CAGGAAGGGG AAAAGATTTC TGCAAAATGAG AATAGCCTGG CAGTCCGTTT
1101 CACCCCTGCC GAAGATGACT CCGGTGACTC CCAGGTTAAG AGTGAGGTTT
1151 AACAGCCTGT CCATCCCAAG CCACTAAGTC CAGATTCAGC AGCCTCCAGT
1201 CTTTCTGAAA GTTCTCCTCC CAAAGCAATG AAGAAGTTTC AGGCACCTGC
1251 AAGAGAGACC TGCGTGAAT GTACAGAGAC AGTCTATCCA ATGGAGCGTC
1301 TCTTGGCCAA CCAGCAGGTG TTTACATCA GCTGCTTCCG TTGCTCCTAT
1351 TGCAACARCA AACTCAGTCT AGGAACATAT GCATCTTTAC ATGGAAGAAT
1401 CTATTGTAAG CCTCACTTCA ATCAACTCTT TAAATCTAAG GGCAACTATG
1451 ATGAAGGCTT TGGGCACAGA CCACACAAGG ATCTATGGGC AAGCAAAAAT
1501 GAAAACGAAG AGATTTTGGA GAGACCAGCC CAGCTTGCAA ATGCAAGGGA
1551 GACCCCTCAC AGCCAGGGG TAGAAGATGC CCCTATTGCT AAGGTGGGTT
1601 TCCTGGCTGC AAGTATGGAA GCCAAGGCCT CCTCTCAGCA GGAGAAGGAA
1651 GACAAGCCAG CTGAAACCAA GAAGCTGAGG ATCGCCTGGC CACCCCCAC
1701 TGAACCTGGA AGTTCAGGAA GTGCCCTTGA GGAAGGGATC AAAATGTCAA
1751 AGCCCAAATG GCCTCCTGAA GACGAAATCA CCAAGCCCGA AGTTCCTGAG
1801 GATGTCGATC TAGATCTGAA GAAGCTAAGA CGATCTCTTT CACTGAAGGA
1851 AAGAGCCGCG CCATTCACTG TAGCAGCTTC ATTTCAAAAG ACCTCTGTCA
1901 AGAGCCCAAA AACTGTCTCC CCACCTATCA GGAAGGCTG GAGCATGTCA
1951 GACCAGACTG AAGAGTCTGT GGCTGGAAGA GTTGCAAAA GGAACAACCT
2001 GCAAAATGCC AAGGCTTCTA AGAAGATGG GAATGCTGGA AAAACAACCT
2051 GCCAAAACAA AGAATCTAAA GGAGAGACAG GGAAGAGAAG TAAGCAAGGT
2101 CATAGTTTGG AGATGGAGAA TGAGATCTTT GTAGAAAATG GTGCAGACTC
2151 CGATGAAGAT GATAACAGCT TCCTCAAACA ACAATCTCCA CAAGAACCCA
2201 AGTCTCTGAA TTGTCGAGT TTTGTAGACA ACACCTTTGC TGAAGAATTG
2251 ACTACTCAGA ATCAGAAATC CCAGGATGTG GAATCTGGG AGGGAGAATG
```

```

2301 GGTCAAAGAG CTCTCTGTGG AAGAACAGAT AAAGAGAAAT CGGTATTATG
2351 ATGAGGATGA GGATCAAGAG TGACAAATTG CAATGATGCT GGGCCTTAAA
2401 TTTCATGTTAG TGTTAGCGAG CCCTGCGCCT TTGTCAAAAT GTGATGCACA
2451 TAAGCAGGTA TCCCAGCATG AAATGTAAAT TACTTGGAAG TAACTTTGGA
2501 AAAGAATTCC TTCTTAAAT CAAAAACAAA AAAAAAAC AAAAAAAC
2551 CATTCTAAT ACTAGAGATA ACTTACTTA ATTCTTCAT TTAGCAGTG
2601 ATCATATCCG TAAGTCTCT AAGGCTTGA ACTGGGAAA TATCCACCT
2651 GATAATAGCC CAGATTCTAC TGTATCCCA AAGGCAATA TTAAGTAGA
2701 TAGATGATTA GTAGTATATT GTTACACACT ATTTTGAAT TAGAGAACAT
2751 ACAGAAGGAA TTTAGGGCT TAAACATTAC GACTGAATGC ACTTTAGTAT
2801 AAAGGGCACA GTTTGTATAT TTTTAAATGA ATACCAATTT AATTTTATG
2851 TATTACCTG TTAAGAGATT ATTTAGTCTT TAAATTTTT AGGTTAATTT
2901 TCTTGCTGTG ATATATATGA GGAATTTACT ACTTTATGTC CTGCTCTCTA
2951 AACTACATCC TGAACTCGAC GTCTGAGGT ATAATACAAC AGAGCACTTT
3001 TTAGGGCAAT TGA AAAACCA ACCTACACTC TTCGGTGCTT AGAGAGATCT
3051 GCTGCTCCC AATAAGCTT TTGTATCTGC CAGTGAATTT ACTGACTCC
3101 AAATGATTGC TTTCTTTCT GGTGATATCT GTGCTTCTCA TAATTACTGA
3151 AAGTGCAAT ATTTAGTAA TACCTTCGGG ATCAGTCTCC CCCATCTTCC
3201 GTGTTAGAGC AAAGTGAAGA GTTTAAAGGA GGAAGAAGAA AGAAGTCTCT
3251 TACACCATT GAGCTCAGAC CTCTAAACCC TGTATTTCCC TTATGATGTC
3301 CCGTTTTTGA GACACTAAT TTTAAATACT TACTAGCTCT GAAATATATT
3351 GATTTTATC ACAGTATTCT CAGGGTGAAA TTAACCAAC TATAGGCCTT
3401 TTTCTTGGGA TGATTTTCTA GTCTTAAGGT TTGGGACAT TATAAAGTCTG
3451 AGTACATTG TTGTACACAG TTGATATCC AAATGTATG GATGGGAGG
3501 AGAGGTGCT TAAGCTGTAG GCTTTTCTT GTACTGCATT TATAGAGATT
3551 TAGCTTAAAT ATTTTATAGA GATGTAAAC ATTCTGCTT CTAGTCTTA
3601 CCTAGTCTGA AACATTTTGA TTCAATAAG ATTTAATTA AAATTGAAA
3651 AAAAAAAAAA AAAA

```

BLAST Results

Entry HS512217 from database EMBL:
human STS SHGC-14654.
Length = 250
Minus Strand HSPs:
Score = 1202 (180.3 bits), Expect = 1.8e-46, P = 1.8e-46
Identities = 242/244 (99%)

Medline entries

95263566:
Three different rearrangements in a single intron truncate
sterol regulatory element binding protein-2 and produce
sterol-resistant phenotype in three cell lines. Role of introns
in protein evolution.

93258417:
Characterization of a pollen-specific cDNA from sunflower
encoding a zinc finger protein.

Peptide information for frame 1

ORF from 94 bp to 2370 bp; peptide length: 759
Category: similarity to known protein

```

1 MESSPFNRRQ WTSLSLRVTA KELS LVNKNK SSAIVEIFSK YQKAAEETNM
51 EKKRSNTENL SQHFRKGLT VLKKKWNPG LGAESHTDSL RNSSTEIRHR
101 ADHPPAEVTS HAASGAKADQ EEQIHPRSL RSPPEALVQG RYPHIKDGED
151 LKDHSTESKK MENCLGESRH EVEKSEISEN TDASGKIEKY NVPLNRLKMM
201 FEKGEPTQTK ILRAQSRAS GRKISENSYS LDDLEIGPGQ LSSSTFDSEK
251 NESRRNLELP RLSETS IKDR MAKYQAAVSK QSSSTNYTNE LKASGGGEIKI
301 HKMEQKENVP PGPEVCITHQ EGEEKISANEN SLAVRSTPAE DSDSDSQVKS
351 EVQQPVHFKP LSPDSRASSL SESSPPKAMK KFGAPARETC VECQKTVYPM
401 ERLLANQQVF HISCFRCSYC NNKLSLGTYA SLHGRIYCKP HFWDLFKSKG
451 NYVDEGFGHRP HKDLWASKNE NEETLERPAQ LANRRETPHS PCVEDAPIAK
501 VGVLAASMEA KASSQQEKED KPAETKKLRI AWPPPTLCS SGSALEEGIK
551 MSKPKWPPED EISKPEVPED VDLDLKKLR SSKLKERSRP FTVAASFQST
601 SVKSPKTVSP PIRKGWSMSE QSEESVGGRV AERKOVENAK ASKKNNGVKG
651 TTWNKESKG ETGKRSKEGH SLEMENENLV ENGADSEDD NSFLLKQSPQ
701 EPKSLNWSF VDNTFAEEFT TQNKSQDVE LWEGEVVKEL SVEEQIKRNR

```

751 YYDEDEDEE

BLASTP hits

Entry CG22818_1 from database TREMBL:
"SREBP-2"; product: "mutant sterol regulatory element binding
protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory
element binding protein-2 (SREBP-2) mRNA, complete cds. Cricetulus
griseus (Chinese hamster)
Length = 839
Score = 1502 (528.7 bits), Expect = 3.9e-154, P = 3.9e-154
Identities = 290/380 (76%), Positives = 322/380 (84%)

Entry S28507 from database PIR:
transcription factor SF3 - common sunflower
Length = 219
Score = 212 (74.6 bits), Expect = 6.3e-18, Sum P(2) = 6.3e-18
Identities = 36/82 (43%), Positives = 55/82 (67%)

Entry NTLIMDOM_1 from database TREMBL:
"SF3"; product: "LIM-domain SF3 protein"; N.tabacum mRNA for
LIM-domain protein Nicotiana tabacum (common tobacco)
Length = 189
Score = 216 (76.0 bits), Expect = 1.0e-16, P = 1.0e-16
Identities = 42/94 (44%), Positives = 57/94 (60%)

Alert BLASTP hits for DKFZphut1_18119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphut1_18119, frame 1

Report for DKFZphut1_18119.1

[LENGTH] 759
[MW] 85225.57
[pI] 6.41
[HOWOL] TREMBL:CG22818_1 gene: "SREBP-2"; product: "mutant sterol regulatory element
binding protein-2"; Cricetulus griseus SRD-2 mutant sterol regulatory element binding protein-
2 (SREBP-2) mRNA, complete cds. 1e-151
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR257w] 3e-05
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGR162w TIF4631 - mRNA cap-binding protein] 1e-04
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGR162w TIF4631 - mRNA
cap-binding protein] 1e-04
[BLOCKS] BL00478B
[PIRKW] zinc finger 9e-16
[PIRKW] DNA binding 9e-16
[SUPFAM] LIM metal-binding repeat homology 9e-16
[PROSITE] MYRISTYL 6
[PROSITE] LIM DOMAIN_1 1
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 4
[PROSITE] CK2_PHOSPHO_SITE 28
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 6
[PFAM] LIM domain containing proteins
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 5.53 %

SEQ MESSPFNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSQYQAAEETNMEKKRSNTENL
SEG
lct1-

SEQ SQHFRKGTLTVLKKKWNPGLAGESHTDSLRNSSTEIRHRADHPPAEVTSAAASGAKADQ
SEG
lct1-

SEQ EEQIHPRSLRSPPEALVQGRYPHIKDGEDLKDHSSTESKKMENCLGESRHEVEKSEISEN
SEG
lct1-

SEQ TDASGKIEKYNVPLNRLKMMFERGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ
SEG


```

1ctl- .....
SEQ  LSSSTFDSEKNESRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI
SEG  .....
1ctl- .....
SEQ  HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKSEVQQPVHFKP
SEG  .....X
1ctl- .....
SEQ  LSPDSRASSLSSESPPKAMKKFQAPARETCVECKTVYPMERLLANQOVFHISCFRCSYC
SEG  XXXXXXXXXXXXXXXX
1ctl- .....ETTTTETTTTCEEEETTEEEETTTTBT
SEQ  NNKLSLGTYSASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEILERPQ
SEG  .....
1ctl- TCBCTTBEEETTEETTTTTCCTTTTCTTT
SEQ  LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQEKEDKPAETKKLRIAWPPPTLG
SEG  .....
1ctl- .....
SEQ  SGALEEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFTVAASFQST
SEG  .....XXXXXXXXXXXXXXXXXXXX
1ctl- .....
SEQ  SVKSPKTVSPPIRKGWSMSEQSEESVGGRAERKQVENAKASKKNGVKTWQNKESKG
SEG  .....
1ctl- .....
SEQ  ETGRKSKEGHSLEMEENENLVENGADSDDDNSFLKQSQPEPKSLNWSSFVDNTFAEEFT
SEG  .....
1ctl- .....
SEQ  TQNKQSQDVELWEGEVVKELSVVEQIKRNRYYDEDEDEE
SEG  .....XXXXXXX
1ctl- .....

```

Prosites for DKFZphut1_18i19.1

PS000001	29->33	ASN_GLYCOSYLATION	PDOC00001
PS000001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS000001	92->96	ASN_GLYCOSYLATION	PDOC00001
PS000001	251->255	ASN_GLYCOSYLATION	PDOC00001
PS000001	286->290	ASN_GLYCOSYLATION	PDOC00001
PS000001	706->710	ASN_GLYCOSYLATION	PDOC00001
PS000004	52->56	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	65->69	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	222->226	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	579->583	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS000005	19->22	PKC_PHOSPHO_SITE	PDOC00005
PS000005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS000005	158->161	PKC_PHOSPHO_SITE	PDOC00005
PS000005	184->187	PKC_PHOSPHO_SITE	PDOC00005
PS000005	220->223	PKC_PHOSPHO_SITE	PDOC00005
PS000005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS000005	253->256	PKC_PHOSPHO_SITE	PDOC00005
PS000005	266->269	PKC_PHOSPHO_SITE	PDOC00005
PS000005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS000005	583->586	PKC_PHOSPHO_SITE	PDOC00005
PS000005	601->604	PKC_PHOSPHO_SITE	PDOC00005
PS000005	604->607	PKC_PHOSPHO_SITE	PDOC00005
PS000005	642->645	PKC_PHOSPHO_SITE	PDOC00005
PS000005	662->665	PKC_PHOSPHO_SITE	PDOC00005
PS000006	19->23	CK2_PHOSPHO_SITE	PDOC00006
PS000006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS000006	55->59	CK2_PHOSPHO_SITE	PDOC00006
PS000006	85->89	CK2_PHOSPHO_SITE	PDOC00006
PS000006	93->97	CK2_PHOSPHO_SITE	PDOC00006
PS000006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS000006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS000006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS000006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS000006	266->270	CK2_PHOSPHO_SITE	PDOC00006
PS000006	294->298	CK2_PHOSPHO_SITE	PDOC00006
PS000006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS000006	326->330	CK2_PHOSPHO_SITE	PDOC00006
PS000006	337->341	CK2_PHOSPHO_SITE	PDOC00006

PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00006	389->393	CK2_PHOSPHO_SITE	PDOC00006
PS00006	467->471	CK2_PHOSPHO_SITE	PDOC00006
PS00006	514->518	CK2_PHOSPHO_SITE	PDOC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS00006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS00006	658->662	CK2_PHOSPHO_SITE	PDOC00006
PS00006	686->690	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	709->713	CK2_PHOSPHO_SITE	PDOC00006
PS00006	714->718	CK2_PHOSPHO_SITE	PDOC00006
PS00006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS00007	223->230	TYR_PHOSPHO_SITE	PDOC00007
PS00007	222->230	TYR_PHOSPHO_SITE	PDOC00007
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	427->433	MYRISTYL	PDOC00008
PS00008	502->508	MYRISTYL	PDOC00008
PS00008	539->545	MYRISTYL	PDOC00008
PS00008	548->554	MYRISTYL	PDOC00008
PS00008	627->633	MYRISTYL	PDOC00008
PS00009	220->224	AMIDATION	PDOC00009
PS00009	662->666	AMIDATION	PDOC00009
PS00478	390->425	LIM_DOMAIN_1	PDOC00382

Pfam for DKFZphut1_18i19.1

HMM_NAME	LIM domain containing proteins		
HMM	*CagCNrpIyDREIvMRAMNKvWHpECFrCcdCqQPLtegdeFYErDGri		
	C	C++++Y+ E++ A+ V+H++CFRC+ C+ L+ G+ + ++ GRI	
Query	390	CVECQKTVYPMERLL-ANQQVFHISCFRCSYCNKLSLGT-YASLHGRI	436
HMM	YCKhDYrrFg*		
	YCK+++ ++F+		
Query	437	YCKPHFNQLFK	447

WO 01/12659

PCT/IB00/01496

DKFZphut1_18i4

group: uterus derived

DKFZphut1_18i4 encodes a novel 220 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of uterus-specific genes.

weak similarity to C.elegans D2085.2

complete cDNA, complete cds, few EST hits

Sequenced by AGOWA

Locus: /map="7q31"

Insert length: 1568 bp

Poly A stretch at pos. 1551, polyadenylation signal at pos. 1523

```
1 GCCGAGCGGA GAGGGTAGAG ACGGGGTTTC ACCGTGTTAG CCAAGATGGT
51 CTCGATCTCC TGACCTCGTG ATCCGCCCGC CTCGGCCTCC CAAAGTGCTG
101 GGATTACAGG CGTGAGCCAC TGCGCCCGGC CTGTTGTACA GTTATTAAAG
151 TTATCATTTA ACATGGAAGA AGATGAGTTC ATTTGAGAAA AAACATTCCA
201 ACGTTATTGT GCAGAATTCA TTAACATTTC ACAACAGATA GGTGATAGTT
251 GGGAAATGGAG ACCATCAAAG GACTGTTCTG ATGGCTACAT GTGCAAAATA
301 CACTTTCAAA TTAAGAAATGG GTCTGTGATG TCACATCTAG GAGCATCTAC
351 CCATGGACAG ACATGTCTTC CCATGGAGGA GGCTTTCGAG CTACCCCTGG
401 ATGATTGTGA AGTGATTGAA ACTGCAGCAG CGTCCGAAGT GATTAAATAT
451 GAGTATCATG TCTTATATTC CTGTAGCTAC CRAAGTGCCTG TACTTTACTT
501 TAGGGCAAGC TTTTTAGATG GGAGACCTTT AACTCTGAAG GACATATGGG
551 AAGGAGTTCA TGAGTGCTAT AAGATGCGAC TGCTACAGGG ACCATGGGAC
601 ACTATTACGC AACAGGAACA TCCAATACTT GGGCAACCCCT TTTTGTACT
651 TCATCCCTGC AAGACGAATG AATTCAATGAC TCCTGTATTA AAGAATTCTC
701 AGAAAATCAA TAAGAATGTC AACTATATCA CATCATGGCT GAGCATTGTA
751 GGGCCAGTTG TTGGGCTGAA TCTACCTCTG AGTATGCCA AAGCAACGTC
801 TCAGGATGAA CGAAATGTCC CTAAACAAGA TTCTCTTCTT GAGTTTAGGA
851 ATTGGGGCAC GAGAATGCC AAGAGTTTAC CTGGCCAGCC CTGGCTTTAA
901 TAGGACTGAT ACCATGGAAT ATTTATCTCT ACCAAGATGT GACATGGATT
951 ATTTTCCCTT TGGACACAAA TGCTACAGC AACTGATGTT TGATAGGCTG
1001 AATGTTTAGA AGAAACACTT CAAAGGGATA CATCATGGCC AGGATGGGTG
1051 GCTCACACCT GTAATCCAAG CACTTTGGGA GGCCAAGGTG GGAGCATCAC
1101 TTGATCCTGG GAGTTCGAGA CCAGCCTGGG CAACATGGTG AAACCCCTGC
1151 GGTACAAAAA AATACAAAAA TTTGCCTGTT TATGGTGGTG TGTTCCTGTA
1201 GTCCAGCTCC CCCAGGAGGC TGAGGTGGGA GGTGGCTTT AACCAGGAG
1251 GCAGAGGTTG CAGTGAGCTG AGACTGTGCC ACTGCAGTCC AGCCTGGGTG
1301 ACAGAGCCAG ACATGTCTCT GGGAAAAAAA AAAAAAATAA AAGACACAT
1351 CACTATAAAT AGCAAAAAAA CAATCTAAC TTATTATAC TAGGAATACC
1401 AACATTATTA GGGCACTTGC AGGTTATTCT TTTCTAGGCC AAGTACTTCA
1451 CTTCATTGTT TCTGACATGG AGATTGAGGG AGAAATGTAT TTGTGTGTTT
1501 ATTTTAATGT AAGATATATA AAAATTAAAT TACTGGATTT ACCTGTCCTT
1551 GAAAAAATAA AAAAAAATAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 1

ORF from 163 bp to 822 bp; peptide length: 220
Category: similarity to unknown protein

PCT/IB00/01496

BLASTP hits

Alert BLASTP hits for DKFZphutel_18i4, frame 1

Pedant information for DKFZphutel_18i4, frame 1

Report for DKFZphutcl_18i4.1

```
SEQ MEDEEFIGEXTGFYRCFAEYFKHSQIQSDWENPWSKSCDGYMKHFQIKNGSVSHLGG
PRD cccccccchhhhhhccccccccccccccccccccccccccccccccccccccccccccccc
SEQ ASTHGQTCLPMEEAFELPLDDCEVETIAAASEVIKYEHVLYSCSYQVPVLYFRASFLDG
PRD cccccccchhhhhhccccccccceehhhhhcchhhhhhhheeeccccccccccccccccccc
SEQ RPLTLCIDIEGWECYHKMRLLGCGPWDTITQENFTLQGFFVLHPCKTNEMFVPLNKSQ
PRD cccccckihhhhhhccccccccccccccccccccccccccccccccccccccccccccccc
SEQ KINKNNVNYITSWLSTVGPPVGLNLPLSYAKATSSQDERNPV
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Prosites for DKFZphut1_18i4.1

(No Pfam data available for DKFZphutel_18i4.1)

DKF2phut1_1811

group: nucleic acid management

DKF2phut1_15j18 encodes a novel 184 amino acid protein with similarity to *S. cerevisiae* putative ribosomal protein YHR148w.

The novel protein is similar to several 40S ribosomal proteins and therefore seems to part of the corresponding ribosome subunit.

The new protein can find application in modulation of ribosome assembly, structure and function.

strong similarity to *S.cerevisiae* YHR148wcomplete cDNA, complete cds, EST hits,
potential start at Bp 45 matches kozak consensus ANNatgG
gene disruption of YHR148w is lethal!

Sequenced by AGOWA

Locus: unknown

Insert length: 1076 bp
Poly A stretch at pos. 1035, polyadenylation signal at pos. 1006

```
1  GCGCGCTCTC AGCTTCGGGT CCTGCGGCTG CGGCTGCGCG CATCATGGTG
51 CGGAAGCTTA AGTTCCACGA GCAGAAGCTG CTGAAGCAGG TGGACTTCCT
101 GAACTGGGAG GTCACCGACC ACAACCTGCA CGAGCTGCGC GTGCTGCGGC
151 GTTACCGGCT GCAGCGGCGG GAGGACTACA CGGCTACAAA CCAGCTGAGC
201 CGTCCCGTGC GTGAGCTGGC GCGCGCGCTG CGCGACCTGC CCGAACGCGA
251 CCAGTTCGCG GTGCGCGCTT CGGCGCGCTG GCTGGACAAAG CTGTATGCTC
301 TCGGCTTGGT GCCACGCGCG GGTTCGCTGG AGCTCTGCGA CTTGTCACG
351 GCCTCTGCTT TCTGCGCGCG CGGCTCTCCC ACCGTGCTCC TCAAGCTGCG
401 CATGGCGCAG CACCTTCAGG CTGCGCTGGC CTTTGTGGAG CAAGGCGACG
451 TACGCGTGGG CCCTGACGTG GTTACCGGAC CGGCTCTCCT TGTACGCGCG
501 AGCATGGAGG ACTTTGTAC TGGGTGGAC TCGTCCAAGA TCAAGCGGCA
551 CGTCTAGAG TACAATGAGG AGCGCGATGA CTTGATCTG GAAGCCTAGC
601 GGATCTCCCA CTTTGCATGG GTGCTTTTGA CAGATGGGAA AACTGAGGCC
651 TGATGCTGGA GATTCTATGA GGTGCTCTC CTCAGGGTGA TCAGACGGTC
701 GTAGGTTCTT AAGAAATTGA TTCAATCAGT GCAGGCCATG CATAGGCCA
751 CGGAGGTGCG GTCCTTGTTC TCCAGGAAT GTTCTTAGAA CTTGACTAC
801 TGATTATTAA TTGACTCTGC CTTGGGAAAC AGTGGGAGT AACTTGGTGC
851 AGCACTGGGG TATTGTTGGA CTGGTTCAAT TCGTTTAACT CGAATTCTTG
901 CTCCTGCGCC TGGTTAAGCT GTGTACAGAT GATGGAGAGT TTGGCCTCAA
951 GTTTTATAAA ACTGAGCGAG ACTAGTGTTC AGGATCTCCT CCCTTGTTTA
1001 AATGTCAATA AATGCCCAA CTGCTTGTGA AGCTCAAAAA AAAAAAAAAA
1051 AAAAAAAAAA AAAAAAAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3
-----ORF from 45 bp to 596 bp; peptide length: 184
Category: strong similarity to known protein

```
1  MVRRLKFHEQ KLLKQVDFLN WEVTDHNLHE LRVLRRLRLQ RREDYTRYNQ
51 LSRVAVRELAR LRLDLPERDQ FRVRASAALL DKLYALGLVP TRGSLELCDF
101 VTASSFCRRR LPTVLLKLRL AOHLQAAVAF VEQGHVRVGP DVVTDPAFLV
151 TRSMEDEFVW VDSSKIKRHV LEYNEERDDF DLEA
```

BLASTP hits

Pedant information for DKFZphute1 1811, frame 3

Report for DKFZphutel_1811.3

```
SEQ      MYVKLKFHEQKLLKQVDLFWNEVTDHNLHLRLVLRRYLRQREDDYTRYNQLSRAVRELAR
SEG
PRD      .XXXXXXXXXXXXX.
        cccchhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ      RLRLDLPERDQFVRASAALLDKLYALGLVPTRGSLCDEFVASSFCRRRLPTVLKLRKM
SEG
PRD      .hhhhhhccccchhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ      AQHLQAQAAVFEQGHVRVGPDDVTPDAFLVTRSMDEFDVTWVDSKIKRHVLEYNEERDDF
SEG
PRD      .hhhhhhhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhhhhhcccccc
SEQ      DLEA
SEG      . . .
PRD      CCCC
```

Prosites for DKFZphute1_1811.3

PS00005	163->166	PKC_PHOSPHO_SITE	PDOC00005
PS00006	153->157	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00007	41->49	TYR_PHOSPHO_SITE	PDOC00007
PS00008	87->93	MYRISTYL	PDOC00008

Pfam for DKFZphutel 1811.3

HMM_NAME	Ribosomal protein S4	
HMM	*MSR.YGPRFKIIRRRPGLPWLtnK.....tklrmrY...LRPqGhWR	
Query	1 MVRKLKLFHEQKLLKQVDLFWVEVTDHNLHELRLVRLRYLRQRREDYTRYN	49
HMM	qRktLsKIRRMsqYrIRLQEKQLKrfYMGNIceRLRLRYrEaEdKKKLD	
Query	50 QLSR--AVREIARRLDLPERDQFVRVASAALLDKILYALGLPV-TRGSLE	96
HMM	YstGenMLIENRLDNIIVFRMGKTLTHHARQLTHHRLIRVNDRIVINP	
Query	97 LCDFTASSFCRRRLPTVLLKLRMAQHLQAIAAFArVEEQGHVRVGPDVTD	146
HMM	SYICRPNDIISIRDkqrMQsHlKwniesPeggmRNHLErNnkkyGeGTIN	

WO 01/12659

PCT/IB00/01496

```

          ++++++ +          ++++++W++ S+          ++R+ + Y+ +
Query    147 AFLVTRS---M-----EDFVTWVDSSK-----IKRHVLEYNEERD 178
HMM      rTIEReWipIkINEILVVEY*
          +++ +
Query    179 DFDLE----- 183
```

DKFZphut1_19f19

group: transmembrane protein

DKFZphut1_19f19 encodes a novel 204 amino acid protein with similarity to murine p24 protein.

Murine p24 is expressed only in brain where it is localized exclusively in neurons. It seems to be a neuron-specific membrane protein localised in intracellular organelles of highly differentiated neural cells and may play a role in the neural organelle transport system. As p24, the novel protein contains 2 transmembrane regions, but it contains not the sequence homologous to the microtubule-binding domain of microtubule-associated proteins present in p24.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to mouse P24 protein ;

membrane regions: 2

Summary DKFZphut1_19f19 encodes a novel 204 amino acid protein, with similarity to mouse P24 protein.

similarity to mouse P24 protein

complete cDNA, complete cds, EST hits,
2 TM-domains

Sequenced by AGOWA

Locus: /map-14.8 cR from top of Chr20 linkage group

Insert length: 2042 bp

Poly A stretch at pos. 1958, polyadenylation signal at pos. 1940

```
1 GCAGGCAGAG AGATGAGGAA ACTGAGACCC AGAAAGGTGG AAGCACTTGT
51 CTAAGGTCAC GCCTCCAGGA AGCAGTGTGT CCACGACTCC AGTCCAAGTG
101 GTCAGGCTCC AGAGCCACAA GTCCGAGGGG TCCATGATGC CGAGCTGCAA
151 TCGTTCCTGC AGCTGCAGCC GCGGCCCAAG CCGGAGGATG GCGAGTGTGT
201 ATGGGGCTCG CTCCTACCTG CACCTCTTCT ATGAGGACTG TCGAGGCACT
251 GCTCTCAGCG ACGACCTCTA GGGACCTCCG GTCTCTGTGC CCGCCGCGCC
301 CTGGCCCTCA CTGTGTGGA AGATCAGCCT GTCTCGGGG ACCTCGCTTC
351 TGCTGCTGGG TGTGGCGGCT CTGACCACTG GCTATGCAGT GCCCCCCAAG
401 CTGGAGGGCA TCGGTGAGGG TGAGTTCCTG GTGTTGATC AGCGGGCAGC
451 CGACTACAAC CAGGCCCTGG GCACCTGTGC CCGGCGAGG ACAGCGCTCT
501 GTGTGGCAGC TGGAGTCTTG CTCGCCATCT GCCTCTTCTG GGCCATGATA
551 GGCTGGCTGA GCCAGGACAC CAAGGCAGAG CCCTTGGACC CCGAAGCCGA
601 CAGCCACGTG GAGGTCTTCG GGGATGAGCC AGAGCAGCAG TTGTACCCCA
651 TTTTCCGCAA TGCCAGTGGC CAGTCATGGT TCTCGCCACC CGCCAGCCCC
701 TTTGGGCAAT CTTCTGTGCA GACTATCCAG CCCAAGAGGG ACTCTGAGCC
751 TGCCACATG GCCTAAGATG TGGGTCTTGG ATCCTTCCCC CTCTCACCA
801 TAACCCCTCT CAGTGTCTTC CCCAATCTCT CCCTTTAGAG CCCAACTCCA
851 GGTCAAATCT GGAGCTCAAA TCCAGTGTCT CCCTCCCCAG GAGTGGGGCC
901 CCAACTCTTC CAAGATACCA GCATTCCTCA AGTCTCCCA AACTTCTTA
951 CCACACCCCT CTTCCCAAGC CCTCAGGGG CAGAAACAT CTCTTCAAC
1001 CCGTCCCCAC TCTCTCTCT GATGACCTT GGGCAACCC TTGCCCTTTC
1051 AAGCCATCAG CTCCTGCCCT TCTGCCATGA GGGCTTTGGA TCAGATTCTT
1101 CTTCTGCCCA GGATGAGGAC ACGCACTGCC CTCATAGAC ACAGATGAAG
1151 GGGTGGGGGT CATTAGCTC GAATGGGTCC CAGATGCTCA CTTGGCCTTT
1201 CCTTGCAGGA TAGGTGAGGA CGTTTGCCTC TCACAGTGTG TCTTCTACCT
1251 GCATTTTGGC ATCAGAGCCC CCCAGCCAC CCACACAGG CAATTACTAG
1301 CCTAGTTGA TAGGTGAGGT GGGTGAAGAA GGCTGGAGGT GACATGTCCG
1351 AGGTACACCA ACAAAGCAGC ATGCAGGAAC TAGAAACACA TCTTCAGCCT
1401 CCTCTGGGCG CAGCTCTTGT GCTACAGGTG GGGCGGAGCC AGCCCTCAC
1451 CTTCTGGGTT CCCTGAGGGT CCTCAGGGTG GAGGACAGGT TTGGCCCAAG
1501 AAGACTAGCC AGAGGCTTGA TGGTCCCAGG TGGCTCTGGA TATACTTTGG
1551 ATATGGATTT AAATGGTCTC TAAGAGCCGG GGGTAGGGGG CAGGAAAGT
1601 GGGTTGCTTT TGCCCTCAA AGTCCACCTA CCTAGAAACC AAGCCACGG
1651 TCTTGGCCGT GACCTGATA ATAAATGGGC TCTCTCAGAG GCGCCAGGCC
1701 CTCCCTCCCC AGCCGGAGGC GTCATCTCTC TTCTGTACCA CTAGAGGGAG
1751 CTCTGATGCA CTGGAGAGC AGCCCTCAAG GCTCTGCCC CTCCCTCCC
1801 TAACCTTAC CTTGAGTCTC CACGAGCCTG AAGGCGCTCC TAGGGATCC
1851 TCAGCGGCCC CCCACAGGCG CACACCTTAC TGCTCTGTGT CCTCAGCCCC
1901 CTTCTCATC CTGCACCCCT TCCATCCAC CTTCCCTTTC AATAACAGC
1951 TGGGATGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2001 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA
```


BLAST Results

Entry HS417348 from database EMBL:
human STS WI-14697.
Length = 290
Minus Strand HSPs:
Score = 1254 (188.2 bits), Expect = 3.0e-50, P = 3.0e-50
Identities = 262/273 (95%)

Medline entries

97334404:
A newly identified membrane protein localized exclusively in
intracellular organelles of neurons.

Peptide information for frame 2

ORF from 134 bp to 745 bp; peptide length: 204
Category: similarity to known protein

1 MFPSCNRSCS CSRGPSVEDG KWGVRSYLH LFYEDCAGTA LSDDPEGPPV
51 LCPRRPWPSL CWKISLSSGT LLLLLGVAAL TTGYAVPPKL EGIGEGEFLV
101 LDQRAADYNQ ALGTCLAGT ALCVAAGVLL AICLFWAMIG WLSQDTKAEP
151 LDPEADSHVE VFGDEPEQQL SPIFRNASGQ SWFSPPASPF GQSSVQTIQP
201 KRDS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19f19, frame 2

TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
complete cds., N = 1, Score = 295, P = 3.8e-26

>TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein,
complete cds.
Length = 196

HSPs:

Score = 295 (44.3 bits), Expect = 3.8e-26, P = 3.8e-26
Identities = 58/139 (41%), Positives = 81/139 (58%)

Query: 2 MPSCNRSCSCSRGPSVEDGK---YGVRSYLHLYEDCAGTALSDDPEGPPVLCPPRPWP 58
M SC+ +C R + +G+ YGVRSYLH FYEDC + + + P R W
Sbjct: 1 MTSCNCTCSRRQAQDTGGYQRYGVRSYLHQFYEDCTASIEWEEDDFQIRSPNR-WS 59
Query: 59 SLCKWISLSSGTL LLLLLGVAALTTGYAVPPKLEIGIGEGEFLVLDQRAADYNQALGTCLLA 118
S+ WK+ L SGT+ ++LG+ L G+ VPPK+E GE +F+V+D A YN AL TC+LA
Sbjct: 60 SVFWKVGILISGTVFVILGLTVLAVGFLVPPKIEAFGEADFMVVDTHAVKYNGALDTCKLA 119
Query: 119 GTALCVAAGVLLAICLFWAM 138
G L G +A CL ++
Sbjct: 120 GAVLFCIGGTS MAGCLLSV 139

Pedant information for DKFZphut1_19f19, frame 2

Report for DKFZphut1_19f19.2

[LENGTH] 204
[MW] 21983.07
[pI] 4.69
[HOMOL] TREMBL:MMP2000_1 product: "P24 protein"; Mouse mRNA for P24 protein, complete
cds. 7e-19
[PROSITE] MYRISTYL 4

WO 01/12659

PCT/IB00/01496

{PROSITE} CAMP_PHOSPHO_SITE 1
{PROSITE} CK2_PHOSPHO_SITE 3
{PROSITE} PKC_PHOSPHO_SITE 1
{PROSITE} ASN_GLYCOSYLATION 2
[KW] TRANSMEMBRANE 2
[KW] LOW_COMPLEXITY 10.29 %

SEQ HMPSCNRSCSCSRGSPVEDGKWYGVRSYLHLFYEDCAGTALSDDPEGPPVLCPRRPWPSL
SEG
PRD ccc
MEMMM

SEQ CWKISLSSGTLTLLLGVAALTTGYAVPPKLEGEGEFVLVDORAADYNOALGTCRLAGT
SEGXXXXXXXXXXXXXXXXXXXXX.....
PRD eeeeecc
MEM MMM

SEQ ALCVAAGVLLAICLFWAMIGWLSQDTKAEPLDPEADSHVEVFGDEPEQQLSPIFRNASGQ
SEG
PRD hhh
MEM MMM

SEQ SWFSPASPFGQSSVQTIQPKRDS
SEG
PRD ccccccccccccccccccccccccc
MEM

Prosite for DKFZphut1_19f19.2

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	176->180	ASN_GLYCOSYLATION	PDOC00001
PS00004	201->205	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00006	16->20	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	92->98	MYRISTYL	PDOC00008
PS00008	119->125	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1_19f19.2)

DKFZphutel_19g19

group: uterus derived

DKFZphutel_19g19 encodes a novel 400 amino acid protein, with strong but partial similarity to a bovine elastin-related protein expressed in fetal calf ligamentum nuchae.

The novel protein contains 2 RGD cell attachment sites.
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

similarity to bovine elastin fragment

complete cDNA, complete cds, EST hits

Sequenced by AGOWA

Locus: map=54.9 cR from top of Chr3 linkage group

Insert length: 3244 bp

Poly A stretch at pos. 3227, polyadenylation signal at pos. 3216

```
1  GTAAGTCCAG  TAAGTCCCGC  TTGGCCCTGG  AGTCCACGCG  GATTTTCGAA
51  GCTGGGGCTG  GCAAGAGGCC  GCTGGACACC  ACGCTCCAGT  CGTCAGCCCA
101  CTTCTAGCT  GAACAGCGCG  AGCGGGCGCG  AGCGAGCCGG  GTCCACCAT
151  GGGCCGGAAT  TATTCCAGTA  CCAGTACCCG  GAGAGAACAT  GTCAAAGTTA
201  AAACCAAGTC  CCAGCCAGGC  TTCTTGAAC  GGCTGAGCGA  GACCTCGGGT
251  GGGATGTTTG  TGGGGCTCAT  GGCCTTCCTG  CTCTCCTTCT  ACCTAATTTT
301  CACCAATGAG  GGCCTGCGAT  TGAAGACGGC  AACCTCATTG  GCTGAGGGGC
351  TCTCGCTTGT  GGTGTCTCCT  GACAGCATCC  ACAGTGTGGC  TCCGAGGAAT
401  GAAGGAAGGC  TGGTGCACAT  CATTGGCGCC  TTACGGACAT  CCAAGCTTTT
451  GTCTGATCCA  AACTATGGGG  TCCATCTTCC  GGCTGTGAAA  CTGCGGAGGC
501  ACGTGGAGAT  GTACCAATGG  GTAGAACTG  AGGAGTCCAG  GGAGTACACC
551  GAGGATGGGC  AGGTGAAGAA  GGAGACGAGG  TATTCCTACA  ACAGTGAATG
601  GAGGTACAGAA  ATCATTCAACA  GCAAAACTTT  CGACCGAGAG  ATTGGCCACA
651  ATAACCCGAG  TSCCATGGCA  GTGAGTCAAT  TCACGGCAAC  AGCCCCCTTT
701  CTCCTAAATTG  GCAGGTTTTT  CTCTCTGTCA  GGCCTCATCG  ACAAGTCGA
751  CAACITCAAG  TCCCTGAGCC  TATCCAAGCT  GGAGGACCCT  CATGTGGACA
801  TCATTGCGCC  TGGAGACTTT  TTCTACCACA  GCGAAAATCC  CAAGTATCCA
851  GAGGTGGGAG  ACTTGGCTGT  CTCTTTTTC  TATGCTGGAC  TGAGCGGCCA
901  TGACCTGAC  CTGGGCCCGAG  CTCACGTGGT  CACTGTGATT  GCCCGGCAGC
951  GGGGTGACCA  GCTAGTCCCA  TTCTCCACCA  AGTCTGGGGA  TACCTTACTG
1001  CTCTGCGACC  ACGGGGACTT  CTCAGCAGAG  GAGGTGTTTC  ATAGAGAAGT
1051  AAGGAGCAAC  TCCATGAAGA  CCTGGGGCCT  GCGGGCAGCT  GGCTGGATGG
1101  CCATGTTTAT  GGGCCTCAAC  CTTATGACAC  GGATCCTCTA  CACCTTGGTG
1151  GACTGGTTTC  CTGTTTTCG  AGACCTGGTC  AACATTGGCC  TGAAAGCCTT
1201  TGCTTCTGT  GTGGCCACCT  CGCTGACCC  GCTGACCGTG  GCGGCTGGCT
1251  GGCTCTTCTA  CCGACCCCTG  TGGGCCCTCC  TCATTGCGGG  CCTGGCCCTT
1301  GTGCCCATCC  TTGTGCTCG  GACACGGGTG  CCAGCCAAAA  AGTTGAGGTG
1351  AAAAGACCC  GGCACCCGCC  CGACACCTGC  GTGAGCCCTA  GGAITCCAGT
1401  CCTCTCTAC  CTCTGACCGA  GCTCATGCC  AGAGCAGAG  CCCCCTCAA
1451  TTTTGACTC  TGCACCCCT  CTCTCTTCA  GGGGCCAGAC  TTGGCAGCAT
1501  GTGCACAGG  TTGGTGTTC  CCAGCTCATG  TCTTCCCCAC  ATCTCTCTT
1551  GCCAGTAAG  AGCTTTGGTG  GGCAGCAGCA  GGCATGAATG  GCAAGCTGAC
1601  AGCTTCTCT  GCTGTTTCT  TCCTCTCTTG  GACTGAGTGG  GTACGGCCAG
1651  CCACCTAGCC  CATTGGCAGC  TGACAACGCA  GACACGCTCT  ACGGAGGCC
1701  GCTGATAAAG  GGCTCAGCCT  TGCCGTGTGC  TGCTTCTCAT  CACTGCACAC
1751  AAGTGCCATG  CTTTGCCACC  ACCACCAAGC  ACATCTGTGA  TCCTGAAGGG
1801  CGGCCGTTAG  TCATTACTGC  TGAGTCTTGG  GTCACCAGCA  GACACACTGG
1851  GCATGGACCC  CTCAAAGCAG  GCACACCCAA  AACACAAGTC  TGTGGCTAGA
1901  ACCTGATGTG  GTGTTTAAAA  GAGAAGAAAC  ACTGAAGATG  TCCTGAGGAG
1951  AAAAGCTGGA  CATATACTGG  GCTTCACTCT  TATCTTATGG  CTGGCAGAAA
2001  TCTTTGTAGT  GTGTGGGATC  TCTGAAGGCC  CTATTTAAGT  TTTTCTCTGT
2051  TACTTTGCTG  CTTTATGTGT  ACTTTCTTAC  CCAAGAGGGA  AGTTTCTTGA
2101  AATAAGATTT  AAAAACAAAA  CAAAAAAAC  ACTTAATATT  TCAGACTGTT
2151  ACAGGAAACA  CCGTTTAGTC  TGTCAGTTGA  ATTCAGAGCA  CTGAAAGGTG
2201  TTAATTTGGG  GTATGTGGTT  TGATTGATTA  AAAGTTACCT  CTGAGTATTT
2251  TGTGTCTCTG  AGAAGCTTTA  CAATGGATGC  TTTTGAACAA  AGTATCAGCA
2301  AAAGGATTTG  TTTTCACTCT  GGGAGGAGAG  GGTGGAGAAA  GCACCTGTCT
2351  TCATCCTCTG  GCATCGGAAA  CTCCCTATG  CACTTGAAGA  TGGTTTAAAA
2401  GATTAAAGAA  ACGATTAAAG  GAAAGGTTG  GAAGCTTTAT  ACTAAATGGG
2451  CTCCTTCTAT  GTGACGCCCC  GTCAACACA  ATCAAGAACT  GAGGCCTGAG
2501  GCTGGTTGTA  CAATGCCACC  GCCTGCCTGG  CTGCTTTCAC  CTGGGAGTGC
2551  TTTGATGTG  GGCACCTGGG  CTTCTAGGG  CTGCTTCTGA  GTGGTTCTTT
2601  CACGTGTTGT  GTCCATAGCT  TTAGTCTTCC  TAAATAAGAT  CCACCCACAC
```

```
2651 CTAAGTCACA GAATTTCTAA GTTCCCCAAC TACTCTCACA CCCTTTTAAA
2701 GATAAAGTAT GTTGTAAACCA GGATGTCTTA AATGATTCTT TGTGTACCTT
2751 TTCTGTCTATA TTCAGAAACC GTTTTGTGCC TGCTGGGAGT AATTCCTTTA
2801 GCAATTAAGT ATTTGGTAGC TGAATAAGGG GTCAGAACTT CTGAAACCAG
2851 AGATCTGTAA TCATCTCTAT TGGCCTGGGG TGCCTGTGCT ATAAATGAGT
2901 TTCTTCACAT GAAAAACACA GCCAGCCCAA GATGACTTAT CTGGGTTTAG
2951 GATTCAATAG TATTCACATA CTGCTTATTA CATGAGCAAT TTCATCAAT
3001 CTCCAACTC TTAAGGATG CTTTCGAAA ACACGCTGTA TACCTAGATG
3051 ATGACTAAAT GCAAAATCCT TGGGCTTTGG TTTTCTTCTA GTAAAGATTT
3101 TAAATAACTG CCGACTTCAR AAGTGTTCTT AAAACGAAAG ATATGTGTTAA
3151 GAAAAATTGG AAGCTTTTGG AAAACCAAT TTGTAATATC ATTGTATTTT
3201 TTATTAAGAG TTTTGTAAATA AATTCTTAAA AAAAAAAAAA AAAA
```

BLAST Results

Entry HS545355 from database EMBL:
human STS WI-14815.
Length = 436
Minus Strand HSPs:
Score = 2040 (306.1 bits), Expect = 6.2e-86, P = 6.2e-86
Identities = 420/426 (98%)

Entry HS932147 from database EMBL:
human STS WI-8531.
Length = 341
Minus Strand HSPs:
Score = 1705 (255.8 bits), Expect = 4.7e-70, P = 4.7e-70
Identities = 341/341 (100%)

Medline entries

86051793:
Bovine elastin cDNA clones: evidence for the occurrence of a
new elastin-related protein in fetal calf ligamentum nuchae.

Peptide information for frame 2

ORF from 149 bp to 1348 bp; peptide length: 400
Category: similarity to known protein

```
1 MAANYSSTST RREHVVKVTS SQPGFLERLS ETSGGMFVGL MAFLLSFYLI
51 FTNEGRALKT ATSLAEGLSL VVSPDSIHVS APENEGRLVH IIGALRTSKL
101 LSDPNYGVHL PAVKLRRHVE MYQWVETES REYTEDGQVK KETRYSINTE
151 WRSEIINSKN FDRKIGHNMF SAMAVESFTA TAPFVQIGRF FLSSGLIDRV
201 DNFKSLSLSK LEDPHVDIIR RGDFFYHSEN PKYPEVGDLR VSFSYAGLSG
251 DDPDLGPAHV VTIARQQRGD QLVPESTKSG DTLLLLHHGD FSAEEVFHRE
301 LRSNSMKTWG LRAAGWAMF MGLNLMTRIL YTLVDWFFVF RDLVNIGLKA
351 FAFCVATSLT LLTVAGWLF YRPLWALLIA GLALVPILVA RTRVPARKLE
```

BLASTP hits

Entry I45887 from database PIR:
elastin - bovine (fragment)
Length = 40
Score = 131 (46.1 bits), Expect = 4.9e-08, P = 4.9e-08
Identities = 31/41 (75%), Positives = 34/41 (82%)

Alert BLASTP hits for DKFZphut19g19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut19g19, frame 2

Report for DKFZphut19g19.2

[LENGTH] 400

```

(MW)          44831.53
(pI)          7.23
(HOMOL)       PIR:I45887 elastin - bovine (fragment) 1e-06
(POSITIVE)    RGD      2
(POSITIVE)    MYRISTYL  3
(POSITIVE)    CAMP_PHOSPHO_SITE  1
(POSITIVE)    CK2_PHOSPHO_SITE   6
(POSITIVE)    TYR_PHOSPHO_SITE   2
(POSITIVE)    PKC_PHOSPHO_SITE   5
(POSITIVE)    ASN_GLYCOSYLATION  1
(KW)          TRANSMEMBRANE  4

SEQ  MAANYSSSTSTREHVVKVTSQPGFLERLSETSGGMFVGLMAFLLSFYLIFTNEGRALKT
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  ATSLAEGSLVVSPOSISHSVAPENEGRLVHIIGALRTSKLLSDPNYGVHLPVKLRHVE
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  MYQWVETESREYTEDGQVKKETRYSYNTEWRSEIINSKNFDREIGHNPNPSAMAVESFTA
PRD  hheehhhhhhecccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....M

SEQ  TAPFVQIGRFLLSSGLIDKVDNFKSLSLKLEDPHVDIIRRGDFFYHSENPKYPEVGDLR
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR.....

SEQ  VSFSYAGLSGDDPDLGPAHVVTVIARQRDQLVPFSTKSGDTLLLLHHGDFSAAEVFHR
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  LRSNSMKTWGLRAAGWAMFMGLNLMTRILYTLVDWFPVFRDLVNIKLKAFACVATSLT
PRD  hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  LLTVAAGWLFYRPLWALLIAGLALVPILVARTVPAKKLE
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR.....

```

Prosites for DKFZphut1_19g19.2

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	9->12	PKC_PHOSPHO_SITE	PDOC00005
PS00005	10->13	PKC_PHOSPHO_SITE	PDOC00005
PS00005	97->100	PKC_PHOSPHO_SITE	PDOC00005
PS00005	276->279	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	332->336	CK2_PHOSPHO_SITE	PDOC00006
PS00007	220->227	TYR_PHOSPHO_SITE	PDOC00007
PS00007	99->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	35->41	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00016	221->224	RGD	PDOC00016
PS00016	268->271	RGD	PDOC00016

(No Pfam data available for DKFZphut1_19g19.2)

DKF2phutel_19g22

group: cell structure and motility

DKF2phutel_19g22 encodes a novel 390 amino acid protein with very strong similarity to tuftelin/enamelin.

Tuftelin/enamelin are matrix proteins of the teeth. As other proteins involved in calcification, these proteins are also expressed in the uterus matrix.

The new protein can find application in modulation of tissue-calcification, especially the uterus.

complete cDNA, complete cds start at Bp 51, EST hits in 3' UTR,
human homolog of mouse tuftelin
tuftelin is described as a matrix protein of teeth but it seems also
to be pressend in the uterus matrix

Sequenced by AGOWA

Locus: unknown

Insert length: 3110 bp

Poly A stretch at pos. 3093, polyadenylation signal at pos. 3071

```
1 GCAGACAGCG GGGTGGACAA GTGGCGTGTG TGCTGCGACC CCGAGGGAAG
51 ATGAACGGGA CGCGGAACGT GTGTACCCTG GTGGACGTGC ACCCAGAGGA
101 CCAGGCGGCG GGCAGCGTGG ACATTCTCAG GCTGACTCTC CAGGCGTAAC
151 TGACAGGAGA TGAACCTTGA CACATAGCCC AGAAGGCGGG CAGGAAGACC
201 TATGCCATGG TGTCCAGCCA CTCAGCTGGT CATTCTCTGG CTTCAAGACT
251 GGTGGAGTCC CATGATGGAC ATGAGGAGAT CATTAAAGTG TACTTGAAGG
301 GGAGGTCTGG AGACAAGATG ATTCACGAGA AGAATATTAA CCAGCTGAAG
351 AGTGAGGTCC AGTACATCCA GGAGGCCAGG AACTGCCTAC AGAAGCTCCG
401 GGAGGATATA AGTAGCAAGC TTGACAGGAA CCTAGGAGAT TCCTCCATC
451 GACAGGAGAT ACAGGTGCTG CTAGAAAGC CAATGGCTT TAGTCAGAGT
501 CCCACAGCCC TGTACAGCAG CCCACCTGAG GTGGACACCT CTATAAATGA
551 GGATGTTGAG AGCTTGAGGA AGACGGTGCA GCATCTGCTG GCCAAGCTTC
601 AGGAGGCCAA GCGGCAACAC CAGTCAGACT GTGTGGCTTT TGAGGTCACA
651 CTCAGCCGGT ACCAGAGGGA AGCAGAACAA AGTAATGTGG CCCTTCAGAG
701 AGAGGAGGAC AGAGTGGAGC AGAAGAGGCC AGAAGTCGGA GAGCTGCAGA
751 GGCCTTGTCT AGGGATGGAG ACGGAGCATC AGGCCTTACT GCGCAAAAGT
801 AGGGAAGGGG AGGTGGCCCT AGAGGAACCT CGGAGCAACA ATGCTGACTG
851 CCAAGCAGAA CGAGAAAGAG CTGCTACCCT GGAAGAGGAA GTGGCCGGGT
901 TGCGGGAGAA GATCCACCAC TTGGATGACA TGCTCAAGAG CCAGCAGCGG
951 AAAGTCCGGC AAATGATAGA GCAGCTCCAG AATTCAAAG CTGTGATCCA
1001 GTCAAAGGAC GCCACCATCC AGGAGCTCAA GGAGAAAATC GCCTATCTGG
1051 AGGCAGAGAA TTTAGAGATG CATGACCCTG TGGAAACCTT GATAGAAAAA
1101 CAAATCAGTC ATGGCAACTT CAGCACCCAG GCCCGGGCCA AGACAGAGAA
1151 CCCGGGCGAG ATTAGGATAT CCAAGCCGCC TACCCGGAAG CCCATGCTGT
1201 TCATCCGAGT GGTGGAACCC TGAGCTGCCT GGAGATGCTT CCTGCCATTG
1251 CTGCTGCCTC TGCTCGGAG AAGCCCACTG CCCCCTTTGG CTGTTAACAC
1301 TGCTTTGAC TTCTGACTGT TCCCCTGGCT GCACCCAGGA CTTCCGGCTC
1351 CTGTGTCTCA CCATTCCCAA GCCCTGGGCC ACTCTAAGCT GGCAGACGG
1401 AGCAGCAGCA CCTATTCAAG GCATCTGAGC CCTTTGAAG ACATTGTCTT
1451 GCAAGCAGGA GCCAGGGCAA TATCTATATT CCTACAGTGA CTATTTTCTT
1501 CTGTAGAGAG CCTCCTTCTT GTTGTAGACT GGACTCTGGC TCGCCCATAA
1551 GCCAGGCCCT CATCAGATTG GGAGAGGTGA CAAGATTTCG CTCAGCCCTA
1601 AAAGCTGGAG ACACAGATGT CCAGAGTGAT TGGAGAATGT CCTGGGGGAA
1651 TGAAGTTCTT TCCACAAACA CAGCTCAGTT CTTAGCAACA AACTGTTTGT
1701 TTTTCTACTT GCTCCATCTG CAGCTACGCT TGCCCTGGCC TCCTGCAGAC
1751 AGATAGTGGG GTTACCTGGC AAGGCCTGGT GAGAGCCAGT GAACCTAAGC
1801 TTTGACTGGG TGGCCTTGTC TTTCTGGGGA GGAGGGAATG TACATTGAGG
1851 GAGTAGCCTT TTGCGGAAAA ATTCTTAGG GCTACAGACA GTCATGTGTG
1901 ACTTCTCTCT GCTGTGAAAA CTCACAGAGT CTCTTTAGGG ATTTTCCCTA
1951 AGGTGTACCA CGAGGCAGAC CTCAGCTTTC TTGACCCAGA GCCTGAAAC
2001 TGTTTTACTT GGGTTCCACC ACTCCAGACA AAATCCTCTT TGTATTTATT
2051 TTGCTAAGTT ATTGGTGCTT TTGCTTACAT CTCATGATTG ATATAATACC
2101 AAAGTTCTAT AGCCTTCTCT TGCAGTATTT GGATTTGCTT GAAACCGGGA
2151 AAATGTTTCC CATTAGGCTT GTTAATGTCA GAGTGACACT ATTATGAATC
2201 TTTCTCTCCC TTTCTCTGCG CTGTTTCTTC TCTCTTCTC CTTCAAACCT
2251 GCTCTGCGAG TAAGGAAGGT GAGTCTACTT TCCCTGAGGC TTTGGGGTCA
2301 GAGTATATGT TGTGAGAGA AAGAGGGCAA TCAGGACTCT TCTGGGACCC
2351 AGATGAGTTC TTCACTAGCC CTTCTGAACC CCTTGCTCCA TAATTGGTCT
2401 TTTATCCTGG CTCTGAATGA CCCTGCAGGT CATCATGGTT TTCTTTTTTT
2451 ATTGTTTTTT TTTTTTTCTG AGACAGAGTC TCACTCTGTC ACCCAGGCTG
2501 GAGTGCAGTG GCGCGATCTC AGCTCACTGC AACCTCTGCG TCCCGGATTT
2551 AAGCGATTCT TGTGCTCAG CTTCCCGAGT AGCTGGGACT ACAGGTGTGC
```

```
2601 CACCACGCCT GGCTGATTTT TGTATTTTGA GTAGAGATGG GGTTCACCA
2651 TACTGGCTAG GCTGGTCTCG AATTCTGTAC CTCAGGTGAT CCACCCACCT
2701 CGGCTTCCCA AAGTGCTAGG ATTATAGGCT TGAGCTACTG TGCCCGGCC
2751 ATGGTGTTTT TCTTTAGGGC TCTTCTTACA GCCTTGAGAA GTAGATAGGC
2801 ATCAGAGTAT GGTACTATAG GAATCAGAAA AATTCAAAAC AAATGTGGAT
2851 TAAGTGTTTA GGCTCTATGT GGCTCAGCA GCCAGAATCC TTAAGTCTGT
2901 GTGTTTCTGT GTCTCAGAC TGGGCTCACA TTCTGGCTTT GTCCATAACA
2951 ATGCTCTGGG ATTTCAAGGA GTTCCCTCAT TTGTAAATG AGGGGTCAG
3001 AGCAGTGAT ATCCATGTTT CTTCCTTTC TGATATTGTT GTCTGTGGCA
3051 TATTCTTGT ATGGCGAATT TAATAAATTA TATTAATGTG TCTAAAAAA
3101 AAAAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

98200312:
Tuftelin--aspects of protein and gene structure

97228909:
Timing of the expression of enamel gene products during mouse tooth development.

91340750:
Sequencing of bovine enamelin ("tuftelin") a novel acidic enamel protein.

Peptide information for frame 3

ORF from 51 bp to 1220 bp: peptide length: 390
Category: strong similarity to known protein

```
1  MNGTRNWCTL VDVHPEDQAA GSVDIILRLTL QGELTGDELE HIAQKAGRKT
51  YAMVSSHAG HSLASELVES HDGHEEIIKV YLGRSGDKM IHEKNINQLK
101 SEVQYIQEAR NCLQKLREDI SSKLDRLGDD SLHRQEIQV LEKPNQFSQS
151 PTALYSSPPE VDTCTINEDVE SLRKTVDLL AKLQEAQRQH QSDCVAFVET
201 LSRVQREAEQ SNVALQREED RVEQKEAEVG ELQRLRLGME TEHQALLAKV
251 REGEVALEEL RSNNAQCQAE REKAATLEKE VAGLREKIHH LDDMLKSQQR
301 KVRQMIEQLQ NSKAVIQSKD ATIQELKEKI AYLEAENLEH HDRMEHLIEK
351 QISHGNFSTQ ARAKTENPGS IRISKPPSPK PMPVIRVVET
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19g22, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1_19g22, frame 3

Report for DKFZphut1_19g22.3

```
[LENGTH] 390
[MW] 44264.09
[pI] 5.68
[MOLEC] TREMBL:AF047704_1 product: "tuftelin"; Mus musculus tuftelin mRNA, complete
cds. 0.0
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w]
2e-11
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 2e-11
[FUNCAT] 1 genome replication, transcription, recombination and repair [M.
jannaschii, MJ1643] 7e-11
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 1e-08
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 6e-08
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGL086w] 6e-08
[FUNCAT] 03.13 meiosis [S. cerevisiae, YNL250w] 7e-08
```

[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 7e-08
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-07
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-07
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 1e-05
 [FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YNL243w] 1e-04
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-04
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 4e-04
 [FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YMR294w] 7e-04
 [EC] 3.6.1.32 Myosin ATPase 8e-09
 [PIRKW] blocked amino end 1e-07
 [PIRKW] nucleus 1e-06
 [PIRKW] citrulline 1e-07
 [PIRKW] tandem repeat 8e-09
 [PIRKW] heterodimer 3e-06
 [PIRKW] DNA repair 2e-06
 [PIRKW] heart 8e-09
 [PIRKW] endocytosis 3e-07
 [PIRKW] transmembrane protein 4e-10
 [PIRKW] zinc finger 3e-07
 [PIRKW] metal binding 3e-07
 [PIRKW] muscle contraction 8e-09
 [PIRKW] acetylated amino end 1e-06
 [PIRKW] actin binding 8e-09
 [PIRKW] microtubule binding 1e-06
 [PIRKW] cell division control 1e-06
 [PIRKW] ATP 8e-09
 [PIRKW] chromosomal protein 3e-06
 [PIRKW] thick filament 8e-09
 [PIRKW] phosphoprotein 1e-145
 [PIRKW] skeletal muscle 8e-09
 [PIRKW] calcium binding 1e-07
 [PIRKW] meiosis 2e-06
 [PIRKW] alternative splicing 7e-08
 [PIRKW] DNA condensation 3e-06
 [PIRKW] coiled coil 4e-10
 [PIRKW] P-loop 8e-09
 [PIRKW] heptad repeat 1e-07
 [PIRKW] methylated amino acid 8e-09
 [PIRKW] immunoglobulin receptor 2e-06
 [PIRKW] peripheral membrane protein 3e-07
 [PIRKW] cardiac muscle 8e-09
 [PIRKW] hydrolase 8e-09
 [PIRKW] muscle 7e-08
 [PIRKW] EF hand 1e-07
 [PIRKW] cytoskeleton 7e-08
 [PIRKW] hair 1e-07
 [PIRKW] smooth muscle 7e-08
 [PIRKW] calmodulin binding 3e-07
 [SUPFAM] conserved hypothetical P115 protein 2e-09
 [SUPFAM] myosin heavy chain 8e-09
 [SUPFAM] RAD50 protein 2e-06
 [SUPFAM] calmodulin repeat homology 1e-07
 [SUPFAM] myosin motor domain homology 8e-09
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-06
 [SUPFAM] tropomyosin 7e-08
 [SUPFAM] protein-tyrosine kinase ret 3e-07
 [SUPFAM] plectin 1e-06
 [SUPFAM] trichohyalin 1e-07
 [SUPFAM] plectestrin repeat homology 2e-06
 [SUPFAM] ribosomal protein S10 homology 1e-06
 [SUPFAM] protein kinase homology 3e-07
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-06
 [SUPFAM] giantin 4e-06
 [SUPFAM] kinesin-related protein KLPA 1e-06
 [SUPFAM] kinesin motor domain homology 1e-06
 [SUPFAM] human early endosome antigen 1 3e-07
 [SUPFAM] M5 protein 2e-06
 [PROSITE] MYRISTYL 1
 [PROSITE] AMIDATION 1
 [PROSITE] CK2_PHOSPHO_SITE 6


```

(PROSITE)      PKC_PHOSPHO_SITE      4
(PROSITE)      ASN_GLYCOSYLATION     2
(KW)           All_Alpha
(KW)           LOW_COMPLEXITY         4.62 %
(KW)           COILED_COIL           35.13 %

SEQ      MNGTRNWCTLVDPVHPEDQAAGSVDIILRLTLQGETGDELEHIAQKAGRKTYAMVSSHSG
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....

SEQ      HSLASELVESHGHEEIKVYLKGRSGDKMIHEKNINQLKSEVQYIQEARNCLQKLREDI
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....

SEQ      SSKLDRLGDSLHRQEIQVLEKPNGFSQSPALYSSPPEVDTCINEDVESLRKTVQDILL
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....cccccccccccccccccccc

SEQ      AKLQEAQRHQSDCVAFEVTLTRYQREAEQSNVALQREEDRVEQKEAEVGELQRRLLGME
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      TEHQALLAKVREGEVALEELRSNNADCAEREKAATLEKEVAGLREKIHHLDDMLKSQQR
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    CC.....cccccccccccccccccccccccccccccccccccccccccccc

SEQ      KVRQMIQQLQNSKAVIQSKDQTIQELKEKIAYLEAENLEMHDRMEHLIEKQISHGNFSTQ
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    ccccccccccc.....cccccccccccccccccccccccccccccccccccccc

SEQ      ARAKTENPGSIRISKPPSPKMPVIRVET
SEG      .....xxxxxxxxxxxxxxxxxxxxx...
PRD      hhcccccccccccccccccccccccccc
COILS    .....

```

Prosite for DKFZphut1_19g22.3

PS00001	2->6	ASN_GLYCOSYLATION	PDOC00001
PS00001	356->360	ASN_GLYCOSYLATION	PDOC00001
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	171->174	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	378->381	PKC_PHOSPHO_SITE	PDOC00005
PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	157->161	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00006	322->326	CK2_PHOSPHO_SITE	PDOC00006
PS00008	355->361	MYRISTYL	PDOC00008
PS00009	46->50	AMIDATION	PDOC00009

(No Pfam data available for DKFZphut1_19g22.3)

DKFZphutel_19h17

group: intracellular transport and trafficking

DKFZphutel_19h17 encodes a novel 879 amino acid protein, with similarity to *N.crassa* osbp oxysterol-binding protein.

The novel protein contains a oxysterol-binding protein family signature. Mammalian oxysterol-binding protein (OSBP) is a protein binds a variety of oxysterols (oxygenated derivatives of cholesterol). OSBP seems to play a complex role in the regulation of sterol metabolism. OSBP is a cytosolic/Golgi receptor for oxysterols such as 25-hydroxycholesterol, and thus a potential target of siphingomyelin turnover and cholesterol mobilization at the plasma membrane and/or Golgi apparatus. Therefore, the new protein seems to be involved in oxysterol metabolism.

The new protein can find application in modulating the response of cells to oxysterols. The protein can be used as marker for the golgi system. The Protein might be used to direct drugs to the golgi system in response to oxidative stress.

strong similarity to *C.elegans* ZK1086.1 and oxysterol-binding proteins

complete cDNA, complete cds, few EST hits
similarity to proteins involved in steroid biosynthesis

Sequenced by AGOWA

Locus: unknown

Insert length: 3828 bp

Poly A stretch at pos. 3811, polyadenylation signal at pos. 3784

```
1  GCGGCGCGCG CCGGCCGGCC CGGAGCACCG AGCTCGCGGC ACGGTAGGAG
51  AAGCCCCCGA GCGCCACAG CATGAAGGAG GAGGCCTTCC TCCGGCGCGG
101 CTTCTCCCTG TGTCCACCTT CCTCCACCCC TCAGAAAGTC GACCCCCCGA
151 AGCTCACCCG GAACCTTGCTC CTCAGCGGAG ACAATGAGCT CTACCCACTC
201 AGCCACGGGA AGGACATGGA GCCCAACGGC CCGTCGCTGC CCAGGGATGA
251 AGGGCCCCCG ACCCAAGCTT CTGCCACGAA GGTGCCACGC GCAGAGTACA
301 GGCTGTGCAA CGGGTCAGAC AAGGAATGTG TGTCCCCCAC CGCAGGGTTC
351 ACCAAGAGGG AGACTCTCAA GGCCGAGAAG GAGAACTACC GCGAGGAGAA
401 GAAGCGCGCC ACACGGCAGC TGCTCAGCGC TCTGACAGAC CCCAGCGTGG
451 TCATCATGGC TGACAGCCTG AAGATCCGGC GCACCCTGAA GAGCTGGACC
501 AAGCTGTGGT GCGTGCTGAA GCGGGGGGTG CTGCTCATCT ACAAGACGCC
551 CAAGGTGGGC CAGTGGGTGG GCACGGTGCT GCTGCACTGC TGCGAGCTCA
601 TCGAGCGGCC CTCCAAGAAG GACGGCTTCT GCTTCAAGCT CTTCCACCCG
651 CTGGATCAGT CCGTCTGGGC CGTGAAGGGC CCCAAGGTTG AGAGCGTGGG
701 CTCCATCACA CAGCCCTTGC CCAGCAGCTA CCTGATCTTC AGGGCCGCCT
751 CCGAGTCAGA TGGTCGCTGC TGGCTGGACG CCCTGGAGCT GGCCCTGCCG
801 TGCTCTAGCC TACTGAGACT GGGCACCTGC AAGCCGGGCC GAGACGGGGA
851 GCCAGGGACC TCGCCAGACG CATCACCTTC ATCGCTCTGT GGGCTGCCAG
901 CCTCAGCCAC TGCCACCCA GACCAAGACC TCTTCCCACT GAACGGGTCT
951 TCCCTGGAGA ACGATGCATT CTCAGACAAG TCGGAGAGAG AGAACCCCTG
1001 GGAGTCAGAT ACCGAGACCC AGGAACATAG CCGGAAGACG GAGAGTGGCA
1051 GCGACCACTC AGAGACCCCT GGGGCCCCGG TCGGAGAGGC GACCACCTAT
1101 GTGGAGCAGG TCCAGGAGGA GCTGGGGGAG CTGGGCGAGG CGTCCCAGGT
1151 GGAGACAGTG TCAGAGGAGA ACAAGAGTCT GATGTGGACC CTGCTGAAGC
1201 AGCTACGGCC AGGCATGGAC CTGTCCCGCG TGGTGCTACC CAGCTTCGTA
1251 CTGGAGCGGC GCTCCTTCCT GAACAAGCTC TCCGACTACT ACTACCACGC
1301 AGACCTGCTC TCCAGGGCTG CGGTGGAGGA GGATGCCTAC AGCCGCAATG
1351 AGCTGGTGCT GCGGTGGTAC CTGTCTGGCT TCTACAAGAA GCCCAAGGGA
1401 ATCAAGAAGC CGTACAACCC CATCTTGGGG GAGACCTTCC GCTGCTGCTG
1451 GTTCCACCCG CAGACTGACA GCGGCACATT CTACATGACA GAGCAGGTGT
1501 CCCACCAACC GCGCGTGTCT GCCTTCCACC TCAGCAACCG GAAGGACGGC
1551 TTCTGATCAA GTGGCAGCAT CACAGCCCAAG TCCAGGTTTT ATGGGAACCTC
1601 GCTCTCGGCG CTGCTGGACG GCAAGGCCAC GCTCACCTTC CTGAACCGAG
1651 CCGAGGATTA CACCCTTACC ATGCCCTACG CCCACTGCAA AGGAATCCTG
1701 TATGGCACGA TGACCCTGGA GCTGGGTGGG AAGGTCACCA TCGAGTGTGC
1751 GAAGAACAAC TTCCAGGCCC AGCTGGAATT CAAACTCAAG CCCTTCTTCG
1801 GGGGTAGCAC CAGCATCAAC CAGATCTCGG GAAAGATCAC GTCCGGAGAG
1851 GAAGTCTCTG CGAGCCTCAG TGGCCTCTGG GACAGGGACG GTTTTATCAA
1901 GGAGGAAGGG AGCGGAAGCA GTGCCCTTTT CTGGACCCCG AGCGGGGAGG
1951 TCCGCAAGCA GAGGCTGAGG CAGCACACGG TGCCGCTGGA GGAGCAGACG
2001 GAGCTGGAGT CCGAGAGGCT CTGGCAGCAC GTCAACAGGG CCATCAGCAA
2051 GGGCGACACC CACAGGGCCA CACAGGAGAA GTTTGCACTG GAGGAGGCAC
2101 AGCGGCAGCG GGCCCGTGAG CGGCAGGAGA GCCTCATGCC CTGGAAGCCG
2151 CAGCTGTTCC ACCTGGAGCC CATCACCCAG GAGTGGCACT ACCGATACGA
2201 GGACCAAGAC CCCTGGGACC CCTGCAAGGA CATGCCCGAG TTGAGCAAG
2251 ACGGGATCTT GCGGACCTTG CAGCAGGAGG CCGTGGCCCG CCAGACCACC
```

```
2301 TTCCTGGGCA GCCAGGGGCC CAGGCACGAG AGGTCTGGCC CAGACCAGCG
2351 GCTTTCGAAG GCCAGCGACC AGCCCTCCGG CCACAGCCAG GCCACGGAGA
2401 GCAGCGGATC CACGCCTGAG TCCTGCCGAG AGCTCTCAGA CGAGGAGCAG
2451 GATGGTGAAT TTGTCCCTGG CGGTGAGAGC CCATGCCCTC GGTGCAGGAA
2501 GGAGGGCGGG CGGCTGCAGG CCCTGCACGA GGCCATCCTC TCCATCCGAG
2551 AGGCCAGGCA GGAGCTGCAC AGGCACCTCT CGGCCATGCT GAGCTCCAGG
2601 GCACGGGCGAG CACAGGCACC GACCCAGGCG CTCTGCAGA GCCCCGATC
2651 CTGGTTCTCT CTCTGCGTGT TCCTGGCGTG TCAGCTGTTC ATTAACACAC
2701 TCCTCAAATA GGAGCCCTGG GGGCAGAGCT CCTGGCCAGT CCCGAGCCCT
2751 CCCTCCAGG CACCCAGCAC TTTAAGCCTG CTCCATGGAG GCAGAGAGGC
2801 CCGGCAAGCA CAGCCACTGT GACCGGGAGT CCAGGCGCAG GAGGACCCG
2851 GGGCCACAAG GCGCTGCGGG CCCAGGTGTG CTGGGCCCCC CTCAGGGGCA
2901 CTGGCCTCTC TGCAGGGCCT TCCGGCCAGC GCTGGCCTTA ATGCTAAAGC
2951 CAAATGCAGC TTCTGCTGTG CGACGCACCT CTGGCCATCT TGCCGTGTCA
3001 CCCCCTGTCC GGCCTCCACT TCCCATGGGG GATGATGGA TTTAGGUTGG
3051 GAGGGCCTGT GGGGGCCCTG CACACTGACA CCCAGCAGC ACTGACTGGG
3101 CAGGTTTGGG GGAGCAGCCA GGGAGCCCCG AGTGGCCAG GAGTCCCCC
3151 ACACACAGAT GCATAGGCCT GCCTCCGGA GACCTGTCC ACATTGCCGG
3201 GACCACCCCT GTGGGGCCAC TGGTGGGTGC CAGGGACAGG TTAGGGCCAC
3251 TCTGGGGAAG GCATTTTGGT TTTTATTCC ACGCTCTGCT GTTTGGATGG
3301 GAGCCCCACA GAGGCAGGTC CTGGAACAC CCCACCCCA CACCTGGACG
3351 CTCGCTCTGG TGGGGGCACA CGCAGGTGGA GGTGGTTGTG GGTGCAGGTG
3401 TGTGCAGGGG TGTGGGGGGC GCAGGGGTGT GGCTTAGCTG GCCCGCACCC
3451 CAGGCCGGGG AGGCTCAAGT TCGCCACTTT ACTCAGACCG ATGCACAGTC
3501 TTCCCATTTT ACACCTTTTT AATAAACATA ATTGAATAT TTAGGTGGG
3551 CTGCGAGCTG CAGTCAGCCT TCACGTCTGG CCTCAGTCCC CGTGTCACTG
3601 CCGCTCTGCG TGTGGGTGTG CGCGTGTGTG AGCCTCTACA CATATATATA
3651 TGTACAGAGC CTAAACAC ATCGTGGCGG TCCGCTCTGA GCTGTAGCGG
3701 GTGGCTTTGT TTCCAGTTT TGTACCCGTG TCCTTGTCTC CCCTCTCTCC
3751 CCATCTGGGG ATGTGTCTGT GTTCCACACC TTGAATATAA CAGACACATA
3801 CGTGTCTCTT TAAAAA AAAA
```

BLAST Results

No BLAST result

Medline entries

98315477:
The pleckstrin homology domain of oxysterol-binding protein recognises a determinant specific to Golgi membranes.

98146266:
A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

98146266:
A Drosophila homologue of oxysterol binding protein (OSBP)--implications for the role of OSBP.

Peptide information for frame 3

ORF from 72 bp to 2708 bp; peptide length: 879
Category: strong similarity to known protein

```
1 MKEEAFLRRR FSLCPPSSPT QKVDPRKLTR NLLSGDNEL YPLSPGKDM
51 PNGPSLPDE GPPTPSSATK VPPAEYRLCN GSKKECVSPT ARVTKKELK
101 AQKENYRQEK KRATRQLLSA LTDPVSUIMA DSLKIRGTLK SWTKLWCVLK
151 PGVLLIYKTP KVGQWVGTVL LHCCELIERP SKKDGFCFKL FHPLDQSVWA
201 VKGPKGESVG SITQPLPSY LIFRAASESD GRCLWDALEL ALRCSLLRL
251 GTCKPGRDGE PGTS PDASPS SLCLPASAT VHPDQDLFPL NGSSLEDAF
301 SDKSERENPE ESDTETQDHS RKTESGSDQS ETPGAPVRRG TTYVEQVEE
351 LGELGEASQV ETVSEENKSL MWTLLKQLRP GMDLSRVVLP TFVLEPRSF
401 NKLSQYIYHA DLLSRAAVEE DAYSRMKLVL RWYLSGFYKK PGYIKKPYNP
451 ILGETFRCCW FHPQDQSRFF YIAEQVSHNP PVSAFVSNR KDFCISGSI
501 TAKSRFYGNS LSALLDGKAT LTLWRAEDY TLTMPYARCK GILYGMTLE
551 LGGKVTIECA KNNFOALEF KLKPFPGGST SINQISGKIT SCEEVLASL
601 GHWRDVFIR EEGSGSSALF WTPSGEVRRQ RLRQHTVPLE EQTELESERL
```

651 WQHYTRAISK GDQHRATQEK FALEEAQROR ARERQESLMP WKPOLFHLDP
701 ITQEMHYRYE DHSPWDPLKD IAOFEQDGL RTLQEAQVAR OTTFLGSPGP
751 RHESGPDQR LKASDQPSG HSQATESGSS TPESCPESLD EQDQGFVPG
801 GESPCPRCRK EARRLQALKE AILSIREAQO ELHRHLSAML SSTARAAQAP
851 TPLGLQSPRS WFLLCVFLAC QLFINHLK

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19h17, frame 3

TREMBL:CEZK1086.2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid
ZK1086, N = 1, Score = 1495, P = 2.7e-153

PIR:S25324 hypothetical protein YKR003w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 574, P = 8.5e-57

TREMBL:CEAF195.7 gene: "C32F10.1"; *Caenorhabditis elegans* cosmid
C32F10., N = 1, Score = 588, P = 8.6e-57

PIR:S46796 hypothetical protein YKR003w homolog YHR001w - yeast
(*Saccharomyces cerevisiae*), N = 1, Score = 585, P = 1.9e-56

TREMBL:NCOSBP.1 gene: "osbP"; product: "oxysterol-binding protein";
N.crassa mRNA for putative oxysterol-binding protein, N = 1, Score =
571, P = 7e-55

TREMBL:AB017026.1 product: "oxysterol-binding protein"; *Mus musculus*
mRNA for oxysterol-binding protein, complete cds., N = 2, Score = 328,
P = 3e-35

>TREMBL:CEZK1086.2 gene: "ZK1086.1"; *Caenorhabditis elegans* cosmid ZK1086
Length = 751

HSPs:

Score = 1495 (224.3 bits), Expect = 2.7e-153, P = 2.7e-153
Identities = 327/663 (49%), Positives = 430/663 (64%)

Query: 129 MADSLKIRGTLKSWTKLWCVLKPGVLLIYKTPKV--GQWVGTVLLHCCELIERPSKKDGF 186
MAD+LKIRG LK W + +CVLKPGL++YK K G WVGTVLL+ CELIERPSKKDGF
Sbjct: 1 MADTLKIRGALKRWNRYYCVLKPGLLILYKHKKADRGDWVGTVLLNHCCELIERPSKKDGF 60

Query: 187 CFKLFHPLDQSVWAVKPGKESVGSIT-QPLPSSYLIFRAASESDGRCLDALELALRCS 245
CFKLFHPLD S+W +GP G+S GS T PL +S+LI RA S+ GRCW+DALEL+ +C+
Sbjct: 61 CFKLFHPLDMSIWGNRGLGQSGFSTLNLPLNTSFLICRAPSDQAGRCWMDALELSFKCT 120

Query: 246 SLLRLGTCKPGRDGEPTSPDASPSSLCGLPASATVHPQODLFFPLNGSSLENDADFSDK-S 304
LL+ T D + G D+S + G + + D D - G A S+ +
Sbjct: 121 GLLKK-TMNE-LDDKNG---DSSMND--GQRDESRMSRSDS-----GDDRELAVSETDA 168

Query: 305 ERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTT---YVEQVOEELGELGEASQVE 361
E+ E D + +DH E G SET +R T ++ +E G G S E
Sbjct: 169 EKHFEIIDDVQDEDH---EDGK-HSETSDT-IREAFTESAWIPSPKEVFGPDG--SLTE 220

Query: 362 TVSEENKSLMWTLTKQLRPGMDLSRVLPFTFVLEPRSFNLKLSDDYYHADLLSRAAVEED 421
V EENKSL+WTLLKQ+RPGMDLS+VVLPTF+LEPRSF KL+DYYHADL+S A E D
Sbjct: 221 EVGEENKSLIWTLLKQIRPGMDLSKVLPFTFILEPRSFLEKLADYYHADLISEAFAEPD 280

Query: 422 AYSRMKLVLRLWYLSGFYKKPKGIKKFPNPLGETFRCCWFHPQDTSRTFYIAEQVSHHPP 481
+ R+ V ++ILSGFYKKPKG+KKFPNPLGETFRCC W HP S TFFYAEQVSHHPP
Sbjct: 281 PFQRIKVKTKFLLSGFYKKPKGLKKFPNPLGETFRCKWEHPD-GSTTFYMAEQVSHHPP 339

Query: 482 VSAFHYSNRKDGFCISGSITAKSRFYGNLSALLDGKATLTLNRAEDYTLTMPYAHCKG 541
VS+ ++NRK GF ISG+I AKS++YGNLSA+L GK LT LN E Y + +PYA+CKG
Sbjct: 340 VSSLFITNRKAGFNISGITLAKSKYYGNLSAILAGKLRLTLNLGETYIYNLPYANCKG 399

Query: 542 ILYGTMTELEGGKVTIECAKNNFQAQLEFKLPFFGGSTSIHQISGKITSGEVSLASLSG 601
I+ GMT+ELGG+V IEC K ++ L+FKLKP GG+ NQI G I G + LAS+ G
Sbjct: 400 IMIGTMTMELGGEVNIIECKTGYRTTLDKFLKPLGGA--YHQIEGSIKYGSDRLASIEG 457

Query: 602 HWRDVFIEKEGSGSSALFWTPSGEVRRQLRQHTVPLEEQTELESERLWQHVTRAISKG 661
WD + IK G W P+ EV + RL ++ + ++EQ E ES +LW+HVT AIS
Sbjct: 458 AWDGVIRIK--GPDGKKELWNPTPEVIKTRLPYIEINMDEQGEWESAKLWRHVTETASNE 515

Query: 662 DQHRATQEKFALEEAQRORARERQESLMPWKPOLFHLDPITQEMHYRYEDHSPWDPLKDI 721
DQ++AT+EK ALE OR RA+ S +P + + F ++ Y + D+ PWD DI
Sbjct: 516 DQYKATEEKTALENDQARAK---SGIPHETKFFKQH-GDDVYIYHAYRPNWNNNDI 570

Query: 722 AQFEQDGLRLTQQEAVAR--QTTFLGSPGPRHRSQPDQLRKASDQPSGHSQATESSG 779
Q E + +++T+ + + + + LGS E S D + + +P + + +
Sbjct: 571 QQIENNYVVKTI SRHSKRKTGNSEQLGSDNTS-EASESDDEVI----EPKIKKKEIVPAK 625
Query: 780 STPESCPESLDE 791
S P + PE++DE
Sbjct: 626 SKPIT-PEVADE 636

Pedant information for DKFZphut1_19h17, frame 3

Report for DKFZphut1_19h17.3

[LENGTH] 879
[MW] 98616.79
[pI] 7.29
[HOMOL] TREMBL:CEZK1086_2 gene: "ZK1086.1"; Caenorhabditis elegans cosmid ZK1086 1e-157
[FUNCAT] 01.06.16 lipid and fatty-acid binding [S. cerevisiae, YHR001w] 3e-55
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YHR001w] 3e-55
[FUNCAT] 3D.03 organization of cytoplasm [S. cerevisiae, YPL145c] 3e-23
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YPL145c] 3e-23
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YAR044w] 5e-20
[BLOCKS] BL00168F
[BLOCKS] BL010130 Oxysterol-binding protein family proteins
[BLOCKS] BL01013C Oxysterol-binding protein family proteins
[BLOCKS] BL01013B Oxysterol-binding protein family proteins
[BLOCKS] BL01013A Oxysterol-binding protein family proteins
[PIRKW] transmembrane protein 1e-19
[SUPFAM] pleckstrin repeat homology 8e-18
[SUPFAM] ankyrin repeat homology 1e-19
[SUPFAM] unassigned ankyrin repeat proteins 1e-19
[PROSITE] MYRISTYL 12
[PROSITE] CAMP_PHOSPHO_SITE 6
[PROSITE] OSBP 1
[PROSITE] CK2_PHOSPHO_SITE 21
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TYR_PHOSPHO_SITE 2
[PROSITE] PKC_PHOSPHO_SITE 20
[PROSITE] ASN_GLYCOSYLATION 3
[PFAM] PH (pleckstrin homology) domain
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 2.96 %
[KW] COILED_COIL 3.53 %

SEQ MKEEAFLLRRRFSLCPPSSTPKVDPKRLTRNLLSGDNELVPLSPGKDMEPNGPSLPRDE
SEG
PRD cccchhhhhhhcc
COILS
MEM
SEQ GPPTPSSATKVPPAEYRLCNGSDKECVSPTARVTKKTLKAQKENYRQEKKRATRQLLSA
SEG
PRD ccc
COILScc
MEM
SEQ LTDPVVIMADSLKIRGTLKSWTKLWCVLKPGLVLLIYKTPKVGQWGTVLLHCCELIERP
SEG
PRD hcc
COILS CCC.....
MEM
SEQ SKRDGFCFKLFHPLDQSVMAVKGPKGESVGSITQPLPSSYLIFRAASESDGRCWLDALEL
SEG
PRD ccc
COILS
MEM
SEQ ALRCSSLLRLGTCKPGRDGEPTSPDASPSSLCGLPASATVHPDQDLFPLNGSSLENDAP
SEG
PRD hhhhhhhhhhhcc
COILS
MEM
SEQ SDKSERENPEESDTETQDHSRKTESGSDQSETPGAPVRRGTTTYVEQVQELGELGEASQV

PCT/IB00/01496

Prosite for DKF2phutel 19h17.3

PS000001	80-~84	ASN_GLYCOSYLATION	PDOC00001
PS000001	29-~295	ASN_GLYCOSYLATION	PDOC00001
PS000001	367-~371	ASN_GLYCOSYLATION	PDOC00001
PS000004	9-~13	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	26-~30	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	95-~99	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	111-~115	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	338-~342	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	762-~766	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	82-~85	PKC_PHOSPHO_SITE	PDOC00005
PS000005	90-~93	PKC_PHOSPHO_SITE	PDOC00005
PS000005	94-~97	PKC_PHOSPHO_SITE	PDOC00005
PS000005	98-~101	PKC_PHOSPHO_SITE	PDOC00005
PS000005	132-~135	PKC_PHOSPHO_SITE	PDOC00005
PS000005	138-~141	PKC_PHOSPHO_SITE	PDOC00005
PS000005	159-~162	PKC_PHOSPHO_SITE	PDOC00005
PS000005	181-~184	PKC_PHOSPHO_SITE	PDOC00005
PS000005	252-~255	PKC_PHOSPHO_SITE	PDOC00005

PS00005	301->304	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS00005	455->458	PKC_PHOSPHO_SITE	PDOC00005
PS00005	488->491	PKC_PHOSPHO_SITE	PDOC00005
PS00005	501->504	PKC_PHOSPHO_SITE	PDOC00005
PS00005	586->589	PKC_PHOSPHO_SITE	PDOC00005
PS00005	647->650	PKC_PHOSPHO_SITE	PDOC00005
PS00005	824->827	PKC_PHOSPHO_SITE	PDOC00005
PS00005	843->846	PKC_PHOSPHO_SITE	PDOC00005
PS00005	857->860	PKC_PHOSPHO_SITE	PDOC00005
PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	227->231	CK2_PHOSPHO_SITE	PDOC00006
PS00006	263->267	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	304->308	CK2_PHOSPHO_SITE	PDOC00006
PS00006	312->316	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00006	362->366	CK2_PHOSPHO_SITE	PDOC00006
PS00006	590->594	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00006	659->663	CK2_PHOSPHO_SITE	PDOC00006
PS00006	713->717	CK2_PHOSPHO_SITE	PDOC00006
PS00006	755->759	CK2_PHOSPHO_SITE	PDOC00006
PS00006	780->784	CK2_PHOSPHO_SITE	PDOC00006
PS00006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS00006	789->793	CK2_PHOSPHO_SITE	PDOC00006
PS00006	824->828	CK2_PHOSPHO_SITE	PDOC00006
PS00007	402->409	TYR_PHOSPHO_SITE	PDOC00007
PS00007	415->424	TYR_PHOSPHO_SITE	PDOC00007
PS00008	137->143	MYRISTYL	PDOC00008
PS00008	163->169	MYRISTYL	PDOC00008
PS00008	274->280	MYRISTYL	PDOC00008
PS00008	326->332	MYRISTYL	PDOC00008
PS00008	381->387	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	541->547	MYRISTYL	PDOC00008
PS00008	552->558	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	613->619	MYRISTYL	PDOC00008
PS00008	728->734	MYRISTYL	PDOC00008
PS00013	860->871	PROKAR_LIPOPROTEIN	PDOC00013
PS01013	474->485	OSBP	PDOC00774

Pfam for DKFZphut1_19h17.3

HMM_NAME	PH (pleckstrin homology) domain		
HMM	*dvIREGWMYKWgswrkstgnWgrRWFvLrndpnrLIYYkddkdekPrYM		
Query	126	VVINADSLKIRGTLKS----WTKLWCVLKP--GVLLIYKTP-KVGQWVG	167
HMM	lIdldcWrMidVEIdWmmdndHCFIiWtrq.....		
Query	168	TVLLHCCCELLIERPSKKD---GFCFKLFHPLDQSVWAVKGPKGESVGSITQ	214
HMMrtYYFQaNeEEMmeWsaIrRaIw*		
Query	215	PLPSSYLIFRAASESDGRCLDALELALR	243

DKF2phutel_19j11

group: uterus derived

DKF2phutel_19j11 encodes a novel 708 amino acid protein with C-terminal similarity to several known proteins, such as human KIAA0231 or murine ras binding protein Sur8.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

Strong similarity to KIAA0231, similarity to ras binding protein Sur8

EST AA854189 extends the sequence (294 Bp), with this sequence complete cDNA,

Sequenced by AGOWA

Locus: unknown

Insert length: 2343 bp

Poly A stretch at pos. 2323, polyadenylation signal at pos. 2295

```
1 GCTCCTGCTA ACCCCATCAG TGTGGAAATG AAAGGCCTGA AGACAGATTT
51 GACCTTCAG CAGTACAGCT TTATAAATCA CATGTGTTAT GAGCGAGCCC
101 TCCACTGGTA TCCCAAGTAT TTCCCTTACC TTGTCCTCAT CCATACCCCTG
151 GTCTTTATGC TCTGCAGTAA CTTTGGTTC AAATTCCTGT GTTCCAGCTC
201 CAAAATAGAA CATTTCATCT CCATTCTGGG GAAGTGTGTT GACTCTCCTT
251 GGACCACAGC GGCTTTATCT GAAGTGTCTG GGGAGGACTC AGAAGAAAAA
301 GACAAACAGGA AGAACAACT GAACAGGTCC AACACCATCC AATCTGGTCC
351 AGAAGGCAGC CTGGTCAACT CTCAGTCTTT AAAGTCCATT CCTGAGAAGT
401 TTGTAGTTGA TAAATCCACT GCAGGGGCTC TGGATAAAAA GGAAGGTGAG
451 CAGGCTAAGG CCTTATTGGA GAAGGTGAAG AAGTTCAGGC TGCATGTGGA
501 AGAAGGTGAT ATTCTATATG CCATGTATGT TCGCCAGACT GTACTTAAAG
551 TTATCAAAAT CTTAATCATC ATTGCATATA ATAGTGTCTC GGTTCCTAAG
601 GTCCAGTTTA CAGTGCAGTG TAATGTGGAC ATTCAAGACA TGACTGGATA
651 TAAAACTTT TCTTGAATC ATACCATGGC ACACCTCTTC TCAAACTGT
701 CCTTTTGCTA TCTGTGCTT GTTAGTATCT ATGGATTGAC GTGCCCTTAT
751 ACCTTATACT GGCCTGTCTA CCGTTCCTTA CGGGAATATT CCTTTGAGTA
801 TGTCCGTCAG GAGACTGGAA TTGATGATAT TCCAGATGTG AAAAATGACT
851 TTGCTTTTAT GCTTCATATG ATAGATCAGT ATGACCTCTC CTATTCCAAG
901 AGATTTGCAG TGTTCCTGTC TGAAGTCAGT GAAACAAAT TAAAGCAGCT
951 GAACTTAAAT AACGAATGGA CTCCTGATAA ACTGAGGCAG AAGCTACAGA
1001 CAAATGCCCA TAAATCGACT GAATTGCCTC TTATCATGCT CTCTGGCCTT
1051 CCAGACACTG TTTTGAATAT CACAGAGTTG CAATCTCTAA AACTTGAAT
1101 CATTAAGAAC GTAATGATAC CAGCCACCAT TGCACGCTA GACAACTTC
1151 AAGAGCTCTC TCTGCACAG TGTCTCTGCA AATCTCACAG TGGCGGCTC
1201 TCTTCTCGA AGGAAACCT CAAGGTCTTG AGCTCAAGT TTGATGACAT
1251 GAGGGAACCT CCCCCCTGGA TGTATGGGCT CCGAAATCTG GAAGAGCTGT
1301 ACCTAGTTGG CTCTCTAAGT CATGATATTT CCAGAAATGT CACCCTTGAG
1351 TCTCTGGGG ATCTCAAAAG CTTTAAATTT CTCTATATCA AAAGCAACGT
1401 TTCCAAAATC CCTCAGGCAG TGGTTGATGT TTCCAGCCAT CTCCAGAAGA
1451 TGTGCATACA TAAATGATGC ACCAAGCTGG TGATGCTCAA CAACTTAAAG
1501 AAGATGACCA ATCTGACAGA GCTGGAGCTG GTCCACTGTG ACCTGGAGCG
1551 TATTCTCAT GCTGTGTCCA GCCTACTCAG CCTCCAGGAA TTGGACCTGA
1601 AGGAAAAACA TCTGAAATCT ATAGAAAGAA TCGTTAGCTT TCAGCACTTA
1651 AGAAAGTTGA CAGTGTGATA ACTGTGGCAT AACAGCATCA CCTACATCCC
1701 AGAGCATATA AGAAATCTCA CCGCCTGGA AGCCTGTCC TTATGTCACA
1751 ATAAATAGA GGTCTGCTT TCCCACCTCT TCCATGCAA CAAGATCCGA
1801 TACTTGGACT TATCGTACAA TGACATTGCA TTTATCCCC CTGAAATGG
1851 AGTTCTACAA AGTTTACAGT ATTTTCCAT CACATGTAAC AAAGTGGAAA
1901 GCCTTCCAGA TGAACCTTAC TTCTGCAAGA AACTTAAAC TCTGAAGATT
1951 GGAAAAACA GCCTATCTGT ACTTTCACCG AAAATTGGAA ATTTGCTATT
2001 TCTTCTCAT TTAGATGTAA AAGGTAATCA CTTTGAATC CTCCCTCCTG
2051 AACTGGGTGA CTGTGGGCT CTGAAGCGAG CTGGTTTAGT TGTAGAAGAT
2101 GCTCTGTTG AAACCTGCC TCTGACGTC CGGGAGCAAA TGAANAACAGA
2151 ATAACTTATT TTTGTTTAAA GTTTGACTGA AACACGCTTC TACCAATAC
2201 AGTATAAATA ATTAGGTAGT CTTAATGCT TTCTATTATT TTTTCTTTT
2251 TCACACAAA TGTACACAAA GATCGCTAA GGAGTATGTA TTTTAAATA
2301 AAATTTAATT GTATTTTTC AATATTAAAA AAAAAAAAAA AAA
```

BLAST Results

No BLAST result

Medline entries

96421675:
Characterization of densin-180, a new brain-specific synaptic protein
of the
O-sialoglycoprotein family.

98337190:
SUR-8, a conserved Ras-binding protein with leucine-rich
repeats, positively regulates Ras-mediated signaling in *C. elegans*.

Peptide information for frame 1

ORF from 28 bp to 2151 bp; peptide length: 708
Category: similarity to known protein
Classification: Cell signaling/communication

```
1 MRGLKTDLDL QQYSEINOMC YERALHWYAK YFPYLVLIHT LVFMLCSNFW
51 FKFPSSSSKI EHFISILGKC FDSFWTTTRAL SEVSGEDSEE KONRKNHNR
101 SWTIQSGPEG SLVNSQSLKS IPEKFWVVKIS TAGALDKKEG EQAKALFEKV
151 KKFRHLVEEG DILYAMYVRQ TVLRVVKFLI IIAVNSALVS KVQFTVDCNV
201 DIQDMTGKYN FSCNHTMAHL FSKLSFCYLC FVSIYGLTCL YTLVWLFYRS
251 LREYSFEYVR QETGDDIDPD VKNDFAFMLH MIDQYDPLYS KRFVFLSEV
301 SENKLEQLNL NNEWTPDKLR QKLQTNANHR LELPLIMLSG LPDVFTEITE
351 LQSLKLEIHK NVMIPTATIAQ LDNLQELSLH QCSVKIHSAA LSFLENLKV
401 LSVKFDDMRE LPPWMYGLRN LEELYLVGSL SHDISRNVTL ESLRDLKSLK
451 ILSIKSNVSK IPQAVVDVSS HLQKMCIHND GTRKLVMLNLI KKMTNLTELE
501 LVHCDLERIP HAVFSLLSLQ ELDLKENNLK SIEEIVSFQH LKRLTVLKLW
551 HNSITYIPEH IKKLTSLERL SFSHNKIEVL PSHLFLCNKI RYLDLSYNDI
601 RFIPPEIGVL QSLQVFSITC NKVESLPDEL YFCKKLKTLK IGKNSLSVLS
651 PKIGNLLFLS YLDVKGHFE ILPPELGDCR ALKRAGLVVE DALFETLPSD
701 VREQMKTE
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_19j11, frame 1

TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene,
partial cds., N = 1, Score = 1408, P = 4.5e-144

TREMBL:AF054827_1 gene: "soc-2"; product: "leucine-rich repeat protein
SOC-2"; *Caenorhabditis elegans* leucine-rich repeat protein SOC-2
(soc-2) mRNA, complete cds., N = 1, Score = 304, P = 5.7e-24

TREMBL:RNU66707_1 product: "densin-180"; *Rattus norvegicus* densin-180
mRNA, complete Cds., N = 1, Score = 311, P = 7.4e-24

TREMBL:AF068921_1 product: "Ras-binding protein SUR-8"; *Mus musculus*
Ras-binding protein SUR-8 mRNA, complete cds., N = 1, Score = 302, P =
1.1e-23

>TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial
cds.

Length = 476

HSPs:

Score = 1408 (211.3 bits), Expect = 4.5e-144, P = 4.5e-144
Identities = 265/471 (56%), Positives = 361/471 (76%)

Query: 237 LTCLYTLWLFYRSLREYSFEYVRQETGDDIDPVKNDFAFMLHIMIDQYDPLYSKRFAVF 296
LT Y+L+W+ SL++YSFE +R+++ DIPDVKNDFAF+LH+ DQYDPLYSKR++F
Sbjct: 1 LTSSSYLWMLRSSLSKQYSFEALREKSNYSIDIPDVKNDFAFILHLADQYDPLYSKRFSIF 60

Query: 297 LSEVSENKLEQLNLNNEWTPDKLRQKLQTNANHRLELPLIMLSGLPOTVFEITELQSLKL 356
LSEVSENKLRQ+NLNNEWTPDKLRQKLQTNANHRLELPLIMLSGLPOTVFEITELQSLKL 356
Sbjct: 61 LSEVSENKLRQINLNNEWTPDKLRQKLQTNANHRLELPLIMLSGLPOTVFEITELQSLKL 120

Query: 357 EIIKNVMIPATIAQLDNLQELSLHQCCKVSIKSAALSFLKENLKVLSVKFDDMRLEPPWY 416
 E+I V +P+ ++QL NL+EL ++ S+ + AL+FL+ENLK+L +KF +M ++P W++
 Sbjct: 121 ELIPEVKLPSAVSQVLNKLKELRVYHSSLVVDHPALAFLEENLKILRLKFTMGKIPRWVF 180

Query: 417 GLRNLEELYLVGSLSHDISRNVTLSELRDLKSLKILSIKSNVSKIPQAVVDVSSHLQKMC 476
 L+NL+ELYL G + + + LE +DLK+L+ L +KS++S+IPQ V D+ LQK+
 Sbjct: 181 HLKRLKELYLSGCVLPQELSTMQLGEGFQDLKWLRTLYLKSSLSRIQVVTDLPLSLQKLS 240

Query: 477 IHNDGTLVLMNLKMTNLTELELVHCDLERIPHAVFSLSLQELDLKENNLKSIEIV 536
 + N+G+KLV+LNNLKRM NL LEL+ CDLERIPH++FSL +L ELDL+ENNLK++EEI+
 Sbjct: 241 LONEGSKLVVLNNLKRMVNLKSLLEISCDLERIPHISFSLNNLHEDLRENNLKTVEEI 300

Query: 537 SFQHLRKLTVLKLWNSITYIPEHIKLTSLERLSFSHNKIEVLPShLFLCNKIRYLDLS 596
 SFQHL+ L+ LKLWNN+I YIP I L++LE+LS HN IE LP LFLC K+ YLDLS
 Sbjct: 301 SFQHLQNLSCCLKLWNNIAYIPAIQIGALSLEQLSDHNNIENPLQLFLCTKLHYLDLS 360

Query: 597 YNDIRFIPPEIGVLSQSYFISITCNKVESLPDELYFCKKLTCLKIGKNSLSVLSFKIGNL 656
 YN + FIP EI L +LQYF++T N +E LPD L+ CKKL+ L +GKNSL LSP +G L
 Sbjct: 361 YNHLTFIPEEIQLSNLQYFAVTNNNIEMLPDGLFQCKKLQCLLKGKNSLNNLSPHVGEL 420

Query: 657 LFLSYLDVKGNNHFEILPELGDCLRALRAGLVVEDALFETLPDVRQMK 707
 L++L++ GN+ E LPPEL C++LKR L+VE+ L TLP V E++T
 Sbjct: 421 SNLTHLELIGNYLETLPELEGCSLKRNLIVEENLLNTLPLPVTRELQT 471

Pedant information for DKFZphut1_19j11, frame 1

Report for DKFZphut1_19j11.1

[LENGTH] 708
 [MW] 81812.82
 [pI] 7.55
 [HOMOL] TREMBL:HSD984_1 gene: "KIAA0231"; Human mRNA for KIAA0231 gene, partial cds.
 1e-149

[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-17
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 3e-09
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YAL021c] 9e-08
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR353c] 3e-07
 [BLOCKS] BL00868f
 [BLOCKS] BL00985B Spermathecal family proteins
 [EC] 3.4.17.3 Lysine carboxypeptidase 1e-08
 [EC] 4.6.1.1 Adenylate cyclase 3e-18
 [PIRKE] blocked amino end 1e-10
 [PIRKE] phosphotransferase 1e-09
 [PIRKE] nucleus 6e-08
 [PIRKE] duplication 3e-18
 [PIRKE] platelet 1e-10
 [PIRKE] tandem repeat 7e-16
 [PIRKE] keratan sulfate 7e-07
 [PIRKE] metallo-carboxypeptidase 1e-08
 [PIRKE] transmembrane protein 1e-10
 [PIRKE] serine/threonine-specific protein kinase 1e-09
 [PIRKE] autophosphorylation 1e-09
 [PIRKE] cartilage 7e-07
 [PIRKE] connective tissue 7e-07
 [PIRKE] magnesium 1e-09
 [PIRKE] cAMP biosynthesis 3e-18
 [PIRKE] ATP 1e-09
 [PIRKE] receptor 1e-09
 [PIRKE] leucine zipper 3e-13
 [PIRKE] glycoprotein 5e-14
 [PIRKE] extracellular matrix 7e-07
 [PIRKE] chondroitin sulfate proteoglycan 7e-07
 [PIRKE] cell adhesion 1e-08
 [PIRKE] hydrolase 1e-08
 [PIRKE] sulfoprotein 7e-07
 [PIRKE] membrane protein 1e-08
 [PIRKE] phosphorus-oxygen lyase 3e-18

[illegible]

(No Prosite data available for DKFZphut1_19j11.1)
(No Pfam data available for DKFZphut1_19j11.1)

DKF2phute1_li2

group: transcription factor

DKF2phute1_li2 encodes a novel 594 amino acid protein similar to signal transducing proteins.

The protein contains 2 WD-40 repeats, which is typical for the beta-transducin subunit of G-proteins. In addition, the protein contains a C3HC4 zinc finger and a leucine zipper. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. Due to the zinc finger the novel protein seems to be a new molecule involved in signal transduction and transcription.

The new protein can find application in modulating/blocking gene expression of genes controlled by this molecule.

similarity to Dictostelium myosin heavy chain kinase

complete cDNA, complete cds, EST hits
[PFAM] Zinc finger, C3HC4 type (RING finger)
[PFAM] WD domain, G-beta repeats
[SCOP] dtbhc_2.46.3.1.1 beta1-subunit of the
signal-transducing G protein 3e-07

Sequenced by BMF2

Locus: /map="16p13.3"

Insert length: 3584 bp

Poly A stretch at pos. 3555, polyadenylation signal at pos. 3537

```
1 GGGCGGGAGG TGCTTCCCAA GGACCGTAGA TGCCTCTCTA GAGCATGAGC
51 TCAGGCAAGA GTGCCCGCTA CAACCGCTTC TCCGGGGGGG CCAGCAATCT
101 TCCACACCCA GACGTCACCA CAGGGACCAG AATGGAAACG ACCTTCGGAC
151 CCGCCTTTTC AGCGTCAACC ACCATCACAA AAGCTGACGG GACCAGCACC
201 TACAAGCAGC ACTGCAGGAC AGCATGCCCC CCATCAGCAC TCCCGCCCGC
251 TCCGACTCCG CCATCTCTGT CCGCTCCCTG CACTCAGAGT CCAGCATGTC
301 TCTCGCTCC ACATTCTGAC TGCCCGAGGA GGAGGAGCAG CCGGAGCCAC
351 TGGTGTGTTG CGAGCAGCCC TCGGTGAAGC TGTGCTGTCA GCTCTGCTGC
401 AGCGTCTTCA AAGACCCCGT GATCACCACG TGTGGGCACA CGTCTGTAG
451 GAGATGCGCC TTGAAGTCAG AGAAGTGTCG CGTGACAAAC GTCAAACTGA
501 CCGTGGTGGT GAACAACATC GCGGTGGCCG AGCAGATCGG GGAGCTCTTC
551 ATCCACTGCC GGCACGGCTG CCGGGTAGCG GGCAGCGGGA AGCCCCCAT
601 CTTTGAGGTG GACCCCGAG GGTGCCCTT CACCATCAAG CTCAGCGCCC
651 GGAAGGACCA CGAGGGCAGC TGTGACTACA GGCCTGTGCG GTGTCCCAAC
701 AACCCAGGCT GCCCCCCGCT GCTCAGGATG AACCTGGAGG CCCACCTCAA
751 GGAGTGGGAG CACATCAAAAT GCCCCACTC CAAGTACGGG TGACGTTCA
801 TCGGAAACCA GGACACTTAC GAGACCCACC TGGAGACTTG CCGCTTCGAG
851 GGCTGAAGG AGTTTCTGCA CGAGACGGAT GACCGTTTCC ACGAGATGCA
901 CGTGGCTCTG GCCCAGAAGG ACCAGGAGAT CGCCTTCTG CGCTCCATGC
951 TGGGAAAGCT CTCGAGAAG ATCGACCAGC TAGAAGAAG CCGTGAGCTC
1001 AAGTTTGACG TCCTGGACGA AAACCAGAGC AAGCTCAGCG AGGACCTCAT
1051 GGAGTTCGCG CGGGACGCAT CCATGTTAAA TGACGAGCTG TCCACATCA
1101 ACGCGCGGCT GAACATGGGC ATCTTAGGCT CCTACGACCC TCAGCAGATC
1151 TTCAAGTGCA AAGGGACCTT TGTGGGCCAC CAGGGCCCTG TGTGGTGTCT
1201 CTGGCTCTAC TCCATGGGTG ACCTGCTCTT CAGTGGGCTC TGTGACAAGA
1251 CCATCAAGGT GTGGACACA TGTACACCTT ACAAGTGTCA GAAGACACTG
1301 GAGGGCCATG ATGGCATCTG GTTGGCTCTC TGCATCCAGG GGTGCNAACT
1351 CTACAGCGGC TGTGAGACT GCACCATCAT TGTGTGGGAC ATCCAGAACC
1401 TGCAGAAGGT GAACACCATC CGGGCCCATG ACAACCCGGT GTGACCGGTG
1451 GTCTCCTCAC ACAACGTGCT CTTACGGGCG TCCTGAAGG CCATCAAGGT
1501 CTGGGACATC GTGGGCACTG AGCTGAAGTT GAAGAAGGAG CTCACAGGCC
1551 TCAACCACTG GGTGCGGGCC CTGGTGGCTG CCCAGAGCTA CCTGTACAGC
1601 GGCTCCTACC AGACAATCAA GATCTGGGAC ATCCGAACCC TTGACTGCAT
1651 CCACGTCTCT GAGACGTCTG GTGGCAGCGT CTACTCCATT GCTGTGACAA
1701 ATCACCACAT TGTCTGTGGC ACCTACGAGA ACCTCATCCA CGTGTGGGAC
1751 ATTGAGTCCA AGGACAGGT GCGGACCTC ACGGCGCACG TGGGACCCGT
1801 GTATGCCCTG GCGGTCTCTT CGACGCCAGA CCAGACCAAA GTCTTCAGTG
1851 CATCTACGA CCGGTCCCTC AGGGTCTGGA GTATGGACAA CATGATCTGC
1901 ACGCAGACCC TGCTGCGTCA CAGGGGAGT GTACCCGCGC TGGCTGTGTC
1951 CCGGGGCGGA CTCTTCTCAG GGGCTGTGGA TAGCACTGTG AAGGTTTGGA
2001 CTTGCTAACA GGATCCAGGC CAGGCTGTGG TTTCCCTGA ACCAGCCCTG
2051 GACCTTTCTG AGCCAGGCTG GCCACATGGG GTGGTCTCGG GGTTCCTGCC
2101 TGCCCCCTGG GCATAGGTGG ACAGGCTCTG GCAGCCGGGC AGTGCCCTCC
2151 CCGTCCCATG CTCGGCGAGC CTCCTCTAC TCGGCACTGT CTTGTCTGCC
2201 CAGCCCCCTCT CTGGGTGCCA GGTACGACGC TTGCCCCGGC CCACCTTCCA
2251 TCCCCACCTT CCAATCCCCAC CTAAGTGA GCGAGGGGCT TTTTACTCCA
2301 CTTTCTTACC GTTTTATAGC TGTATGTAGA TTTGGTACC TCTGTTTGA
```

```
2351 AATAAATGCT CCACAGACTG TGGCTGTGAG TGGGGACAGC TCCTCGGGAC
2401 AAGGGGGCTG TGTGTGGCCT TGAGGTTGGT GTGCACAGGC ACTGGCTGCT
2451 GTGAGTGGGG GGGCATGGGG CAGTTTCCTT TGGTGGACCC CAGGACTTCG
2501 GCCCACTCCG GGGCCTCCCC TCCCTGCTAG GAGGCAACTC GTACACCCCA
2551 AGCTGCTGGC CTCAGTCCC ATCTCCCCCA ACACATGTGC CCCCAAAAG
2601 TGAGCCAGGC ACCTCTGTTT CCTGCTGTTT ATTGACAGCC GACGGCAGCG
2651 CCTTGCCAG ACCTCCCTTG CCCACCTGCT GGAGCCAGC CTGTGCCGCC
2701 CTCTGAGGAG AGGCTGGGG GGACAGCTGG GCAGTCCAC TCGCAGGGA
2751 ACAGGGGGTG AGACAGCAGG AAGGGGCCCT GCACGCGGG AGGCACATC
2801 CGCAGCGGC CTCACCGGC CCACACCCAC AATCGCTGCT TTTCGGCAT
2851 TTTTAAATTT TTTTAAAG AAALGTCAA GTTGTGCCA ACACGTGGA
2901 TCAGCAAACA CGATAGAGGA GACCAGTCA TACTTCTTG AGGGGCGAGG
2951 AGGAGAGAGG AAAAGGGAGG GCGAGAATGA CCACACAACA CAGCCTTGA
3001 CCATGAGCAG AAGCGTCCGT GGGAACTCCA CTGGGGTGA TGGCTGCCT
3051 GCACAGCCCC TGGAGAGGGG GCCAGGCACA CCCTCAGAG AGCTGCAAGC
3101 CGTGGCCTG GCCTGTACA TGCCCTGCTT CCACGTGGCT GCCACGCTGA
3151 CACACCCACA TTCACCAAC CCACCCGGC CCTGGGAGC AGCCACGCA
3201 GGAGGAGGAC ACGGCCCGG AGAGCAAGGC ACAACCTGA GTTCTGGGG
3251 CGCAGAGAAC TTAGGAGAGA AGCAGGAGG AGCCCCGGC AGAGCACCCG
3301 CCCCCGGGCC CCAGCCTTCC ACCTGTGCTA GCAGCCTGG GCTCCACTC
3351 TGGCCGGAGG AAGAGCCGCA GGCAGACAGC CTGGGCTCT AACAGCTTT
3401 GTCCGGAGCT AGACTTCGTG TCCTTTCAGT TGGTAAATGG TTTTCTATG
3451 AATCAATAAT ATTTCTTCTT TTAATATAT ATTTGTAAA GTTATACCTT
3501 TTTGTTTCTC TGGGGAAATC CGCTCAGCT CATTCCCAAT AAATTAATAC
3551 TCTGATAAA AAAAAAAAAA AGAAAAAAAA AAAA
```

BLAST Results

Entry HSBE from database EMBL:
Homo sapiens (clone exon trap d5) chromosome 16p13.3 gene, exon.
Score = 2375, P = 7.1e-101, identities = 475/475

Entry HSBD from database EMBL:
Homo sapiens (clone exon trap d32) chromosome 16p13.3 gene, exon.
Score = 876, P = 3.0e-31, identities = 176/177

Medline entries

95122486:
Structural analysis of myosin heavy chain kinase A from Dictyostelium. Evidence for a highly divergent protein kinase domain, an amino-terminal coiled-coil domain, and a domain homologous to the beta-subunit of heterotrimeric G proteins.

96149460:
Dictyostelium myosin heavy chain kinase A regulates myosin localization during growth and development.

97277316:
Identification of a protein kinase from Dictyostelium with homology to the novel catalytic domain of myosin heavy chain kinase A.

96009891:
A gene responsible for vegetative incompatibility in the fungus Podospora anserina encodes a protein with a GTP-binding motif and G beta homologous domain.

Peptide information for frame 2

ORF from 224 bp to 2005 bp; peptide length: 594
Category: similarity to known protein
Prosite motifs: ZINC_FINGER_C3HC4 (70-80)
LEUCINE_ZIPPER (436-458)
LEUCINE_ZIPPER (436-458)
G_BETA_REPEATS (335-355)
G_BETA_REPEATS (376-391)

1 MPPISTPRRS DSAISVRSIH SESSMSLRST FSLPEEEEP EPLVFAEQPS
51 VKLCCQLCCS VFKDPVITTC GHTFCRRAL KSEKCPVDNV KLTVVVNNIA
101 VAEQIGELFI HCRHGCRVAG SGKPPIFEVD PRGCPPTIKL SARKOHEGSC
151 DYRPVRCFNN PSCPPLLRNN LEAHLKECEH IKCPHSKYGC TFIGNQDTYE
201 THLETCRFEG LKEFLQQTDD RFHEMHVALA QKQDEIAFLR SMLGKLSEKI
251 DQLEKSLELK FQVLDENQSK LSEDLMEFRR DASMDELNDEL SHINARLNMGSI
301 LGSYDPQQIF KCKGTFVGHQ GPVWCLCVYS MGDLLFSGSS DKTIKVMDTC
351 TTYKCKTLE GHDGIVLALC IQGCKLYSGS ADCTIIVMDI QNLQKVNTIR
401 AHDNPVCTLV SSHNVLFSGS LKAIKVMDIV GTTELKRLREL TGLNHWVRAL
451 VAAQSYLYSG SYQTIKIWDI RTLDCHIHLQ TSGGSVYSIA VTNHHIVCGT
501 YENLIHVMDI ESKEQVRLT GHVGTVALA VISTPDQKV PSASYDRSLR
551 VWSMDNMICT QTLRHQGSV TALAVSRGR LFSGAVDSTVK VWTG

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phute1_i12, frame 2

SWISSPROT:KMH8_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B), N = 1, Score = 419, P = 3.6e-37

SWISSPROT:HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1., N = 1, Score = 392, P = 3.1e-33

SWISSPROT:YDJ5_SCHPO HYPOTHETICAL 67.1 KD TRP-ASP REPEATS CONTAINING PROTEIN C57A10.05C IN CHROMOSOME I., N = 1, Score = 357, P = 4.1e-30

TREMBL:AF032878.1 gene: "slimb"; product: "Slimb"; Drosophila melanogaster Slimb (slimb) mRNA, complete cds., N = 1, Score = 347, P = 1.7e-29

>SWISSPROT:KMH8_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B).
Length = 732

HSPs:

Score = 419 (62.9 bits), Expect = 3.6e-37, P = 3.6e-37
Identities = 96/268 (35%), Positives = 158/268 (58%)

Query: 325 CLCVYSMGDLLFSGSSDKTIKVMWD-TCTTYKCKQKTELEHGDGIVLALCIQGCKLYSGSADC 383
C+C +LLF+G SD +I+V+D +C +TL+GH+G V ++C L+SGS+D
Sbjct: 467 CIC----DNLLFTGCDNSIRVYDYKSNMECVQTLKGHEGPVESICYNDQYLFSGSSDH 522

Query: 384 TIIVMDIQNLQKVNTIRAHNDNPVCTLVSSHNVLFSGSL-KAIKVMWDIVGTTELKRLKELTG 442
+I VMD++ L+ +T+ HD PV T++ + LFSGS K IKVMD+ L+ K L
Sbjct: 523 SIKVMWLKLRICFTLEGHDKPVHTVLLNDKYLFGSSDKTIKVMWL--RTLECKYLTLES 580

Query: 443 LNHVVRALVAAQSYLYSGSY-QTIKIWDI RTLDCHIHLQTSGGSVYSIAVTHHHIVCGTY 501
V+ L + YL+SGS +TIK+WD++T C +L+ V +I + ++ G+Y
Sbjct: 581 HARAVKTLICISGQYLFSGSNDKTIKVMWLKTRFCNYTLKGHTKWTTCILGTNLVSGSY 640

Query: 502 ENLIHVWDIESKEQVRLTGHVGTVALAVISTPDQTKVFSASYDRSLRVWSMDNMICTQ 561
+ I VW++S E TL GH V + + D+ +F+AS D +++++W ++ + C
Sbjct: 641 DKTIRVWNLKSLECSATLRGHRWVEHVMVIC--DKL-LFTASDDNTIKIWDLETLCMT 696

Query: 562 TLLRHQGSV TALAVSRGR--LFSGAVDSTVKVW 592
TL H +V LAV + + S + D ++VM
Sbjct: 697 TLEGHNATVQCLAVMEDKKCVISCSHDQSIRVW 729

Score = 415 (62.3 bits), Expect = 1.2e-36, P = 1.2e-36
Identities = 113/303 (37%), Positives = 166/303 (54%)

Query: 255 KSLEL-KFDVLDENQSKLSEDLMEFRDASMLNDEL-SHINARLNMGILGS-----YD 305
KS++L K ++L N+ K S +L + ++ + SH+ N+ G YD
Sbjct: 427 KSIDLEKPEILINKKKESINLETIKLIETIKGYHTSHLCICDNLLFTGCDNSIRVYD 486

Query: 306 -PQIIFKCKGTTFVGHQGPVWCLCVYSMGDLLFSGSSDKTIKVMWDCTTYKCKQKTELEHGDG 364
Q +C T GH+GPV +C Y+ LFSGSSD +IKVMD +C TLEHGD
Sbjct: 487 YKSQNMCEVQTLKGHEGPVESIC-YN-DQYLFSGSSDHSIKVMDL-KKLRCIFTLEGHDK 543

Query: 365 IVLALCIQGCKLYSGSADCTIIVMDIQNLQKVNTIRAHNDNPVCTLVSSHNVLFSGSL-KA 423
V + + L+SGS+D TI VMD++ L+ T+ +H V TL S LFSGS K
Sbjct: 544 PVHTVLLNDKYLFGSSDKTIKVMWLKTLCKYLTLESHARAVKTLICISGQYLFSGSNDKT 603

Query: 424 IKVMDIVGTTELKRLKELTGLNHWVRALVAAQSYLYSGSY-QTIKIWDI RTLDCHIHLQTS 482
IKVMD+ + L G WV + + LYSGSY +TI+M++++L+C L+
Sbjct: 604 IKVMDL--RTFRCNLYLKGHTKWTTCILGTNLVSGSYDKTIRVWNLKSLECSATLRGH 661

Query: 483 GGSVYSIAVTHHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVYALAVISTPDQTKVFS 542
 V + + + + + N I +WD+E+ TL GH TV LAV D+ V S
 Sbjct: 662 DRWVEHNVICDKLLFTASDDNTIKIWDLTLRCNTTLEGHNATVQCLAVWE--DKKCVIS 719

Query: 543 ASYDRSLRW 552
 S+D+S+RVW
 Sbjct: 720 CSHDQSIRW 729

Score = 262 (39.3 bits), Expect = 3.2e-19, P = 3.2e-19
 Identities = 60/184 (32%), Positives = 109/184 (59%)

Query: 352 TYKCKTLECHDGIIVLALCIQGCKLYSGSADCTIIVWDI--QNLQKVNTIRAHDPVCTL 409
 T R +T+G+ + LCI L+G +D +I V+D QN++ V T++ H+ PV ++
 Sbjct: 450 TIKLIETIKGYH-VTSHLCIDNLLFTGCSNIRVYDYKSONMECVQTLKGHEGPVESI 508

Query: 410 VSSHNVLFGSGLK-AIKVWDIVGTLELKKELTGLNHWVRALVAAQSYLYSGSY-QTIKI 467
 + LFGS +IKVWD+ +L+ L G + V ++ YL+SGS +TIK+
 Sbjct: 509 CYNDQYLFGSSSDHSIKVWDL--KKLRCIFLLEGHDKPVHTVLLNDKYLFGSSSDKTIKV 566

Query: 468 WDITLDCIHVLQTSGGSVYSIAVTHHHIVCGTYENLIHVWDIESKEQVRLTGHVGTVY 527
 WD++TL+C + L++ +V ++ ++ G+ + I VWD+++ TL GH V
 Sbjct: 567 WDLTLECKYLTESHARAVKTLICISQYLFSGSNOKTIKWDLKTRCNVTLKGHTKWT 626

Query: 528 ALAVIST 534
 + ++ T
 Sbjct: 627 TICILGT 633

Score = 173 (26.0 bits), Expect = 1.7e-09, P = 1.7e-09
 Identities = 43/118 (36%), Positives = 65/118 (55%)

Query: 310 FKCKGTFVGHQGPVWCLCVYSGDLFGSSDKTIKWDTCTTYKCKTLECHDGIIVLAL 369
 F+C T GH V +C+ +G L+SGS DKTI+VW+ + +C TL GH V +
 Sbjct: 612 FRCNVTLKGHTKMTTICI--LGTNLYSGSYDKTIRVWNL-KSLECSATLRGHDRWVEHM 668

Query: 370 CIQCKLYSGSADCTIIVWDIQLQKVNTIRAHDPV-CTLVSSH--VLFGSLKAIKV 426
 I L++ S D TI +WD++ L+ T+ H+ V C V V+ ++I+V
 Sbjct: 669 VICDKLLFTASDDNTIKIWDLTLRCNTTLEGHNATVQCLAVWEDKKCVISCSHDQSIRV 728

Query: 427 W 427
 W
 Sbjct: 729 W 729

Pedant information for DKF2phute1_1i2, frame 2

Report for DKF2phute1_1i2.2

[LENGTH] 594
 [MW] 66541.94
 [pI] 6.64
 [HOMOL] SWISSPROT:KMHB_DICDI MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B). 3e-37

[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 01.01.04 regulation of amino-acid metabolism [S. cerevisiae, YIL046w] 5e-21
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YCR072c beta-transducin family] 2e-15
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YFL009w] 1e-14
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 1e-13
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 2e-11
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 3e-11
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 8e-09
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 2e-07
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 5e-07
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 5e-07

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 3e-06
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YKL021c] 2e-04
 [FUNCAT] 01.03.07 deoxyribonucleotide metabolism [S. cerevisiae, YOR269w] 2e-04
 [FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YOR212w] 0.001
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YOR212w] 0.001
 [BLOCKS] BL00678
 [BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
 [SCOP] dltbpd 2.46.3.1.1 betal-subunit of the signal-transducing 3e-10
 [EC] 2.7.1.129 Myosin-heavy-chain kinase 3e-26
 [PIRKEW] phosphotransferase 3e-26
 [PIRKEW] nucleus 1e-06
 [PIRKEW] plasma 9e-08
 [PIRKEW] duplication 3e-25
 [PIRKEW] hormone 9e-08
 [PIRKEW] zinc 3e-09
 [PIRKEW] cell cycle control 4e-13
 [PIRKEW] transmembrane protein 3e-12
 [PIRKEW] zinc finger 1e-08
 [PIRKEW] stomach 9e-08
 [PIRKEW] DNA binding 9e-06
 [PIRKEW] autophosphorylation 3e-26
 [PIRKEW] phosphoprotein 3e-26
 [PIRKEW] signal transduction 5e-08
 [PIRKEW] heterotrimer 5e-08
 [PIRKEW] coiled coil 3e-26
 [PIRKEW] multimer 3e-26
 [PIRKEW] transcription regulation 4e-10
 [PIRKEW] GTP binding 5e-08
 [SUPFAM] chromobox homology 9e-06
 [SUPFAM] RING finger homology 3e-09
 [SUPFAM] coatomer complex beta' chain 1e-07
 [SUPFAM] WD repeat homology 3e-26
 [SUPFAM] yeast coatomer complex alpha chain 3e-12
 [SUPFAM] GTP-binding regulatory protein beta chain 5e-08
 [SUPFAM] PRL1 protein 2e-09
 [PROSITE] WD_REPEATS 2
 [PROSITE] LEUCINE_ZIPPER 1
 [PROSITE] MYRISTYL 14
 [PROSITE] CK2_PHOSPHO_SITE 4
 [PROSITE] ZINC_FINGER_C3HC4 1
 [PROSITE] PKC_PHOSPHO_SITE 18
 [PROSITE] ASN_GLYCOSYLATION 1
 [PFAM] Zinc finger, C3HC4 type (RING finger)
 [PFAM] WD domain, G-beta repeats
 [KW] Irregular
 [KW] 3D
 [KW] LOW_COMPLEXITY 6.23 %
 [KW] COILED_COIL 6.73 %

SEQ MPPISTPRRSDSATSVRSLHSESSMSLRSTFSLPEEEEEPEPLVFAEQPSVKLCCQLCCS
 SEGXXXXXXXXXXXXXXXXXXXXX.....XXXXXXXXXX.....
 COILS
 1gg2B
 SEQ VFKDPVITTCGHTFCRRALKSEKCPVDNVKLTVVVNNIAVAEQIGELFIHCRHGCRVAG
 SEG
 COILS
 1gg2B
 SEQ SGKPPIFEVDPRGCPFTIKLSARKDHEGSCDYRPVRCPPNNPSCPPLLRMNLEAHLKECEH
 SEG
 COILS
 1gg2B
 SEQ IKCPHSKYGCTFIGNQDTYETHLETCTFEGLKEFLQOTDDRFHEMHHVALAQKDQEI AFLR
 SEG
 COILSCCCCCCCCCCCCCCCC
 1gg2B
 SEQ SMLGKLSEKIDOLEKSLKFDVLDENQSKLSEDLMEFRDASMLNDELSHINARLNMGI
 SEG
 COILS CCCCCCCCCCCCCCCCCCCCCCCCCC.....
 1gg2B
 SEQ LGSYDPQQIFKCKGTFVGHQGPVWCLVYSMDLFFSGSSDKTIKWDCTTTYKCKQKTLK
 SEG
 COILS
 1gg2BEECCCCCEEEETTTTCEEEETTTTBBBEEG-GGCEEEEEE


```
SEQ      GHDGIVLALCIQCKLYSGSADCTIIWVDIQNLQKVNTIRAHDPVCTLVSSHNVLFSGS
SEG      .....
COILS    CCCCCEEEEETTCEEEEEETTCEEEEETTTTTEEEEE-CTTTTCCCEE.....
lqq2B

SEQ      LKAIKVMDIVGTELKLLKELTGLNHWVRALVAAQSYLYSGSYQTIKIWDIRTLDCIHVLQ
SEG      .....XXXXXXXXXXXXX.....
COILS    .....
lqq2B

SEQ      TSGGSVYSIAVTNHHIVCGTYENLIHWMDIESKEQVRTLTGHVGTVYALAVISTPDQTKV
SEG      .....
COILS    .....
lqq2B

SEQ      FSASYDRSLRVMSMDNMICTQTLLRHQGSVTALAVSRGLFSGAVDSTVKVWTC
SEG      .....
COILS    .....
lqq2B    .....
```

Prosites for DKF2phut1_i12.2

PS00001	267->271	ASN_GLYCOSYLATION	PDOC00001
PS00005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS00005	15->18	PKC_PHOSPHO_SITE	PDOC00005
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	50->53	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	121->124	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	141->144	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	340->343	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	464->467	PKC_PHOSPHO_SITE	PDOC00005
PS00005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	330->334	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00008	115->121	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	194->200	MYRISTYL	PDOC00008
PS00008	299->305	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	364->370	MYRISTYL	PDOC00008
PS00008	379->385	MYRISTYL	PDOC00008
PS00008	419->425	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	484->490	MYRISTYL	PDOC00008
PS00008	499->505	MYRISTYL	PDOC00008
PS00008	524->530	MYRISTYL	PDOC00008
PS00008	568->574	MYRISTYL	PDOC00008
PS00008	583->589	MYRISTYL	PDOC00008
PS00518	70->80	ZINC_FINGER_C3HC4	PDOC00449
PS00029	436->458	LEUCINE_ZIPPER	PDOC00029
PS00678	335->350	WD_REPEATS	PDOC00574
PS00678	376->391	WD_REPEATS	PDOC00574

Pfam for DKF2phut1_i12.2

```
HMM_NAME      WD domain, G-beta repeats
HMM            *MrGHnnWVVCVaFSPDGrWFIvSGSWDgTCRLWD*
               ++GH ++VMC+ +  G + ++SGS D+T++WD
Query          316  FVGHQGPVWCLCVYSMGDL-LFSGSSDKTIKVWD      348

22.93   519   553   1   34 dkf2phut1_i12.2 similarity to Dictostelium myosin heavy chain
kinase
Alignment to HMM consensus:
```

WO 01/12659

PCT/IB00/01496

Query *MrGHnnWVWCVaF..SPDGzWFIvSGSWDgTCRLND*
++CH ++V++++A+ +PD ++S+S D+++R+W+
dkfzphutel 519 LTGHVGTVYALAVISTPDQTK-VFSASYDRSLRVWS 553

HMM_NAME Zinc finger, C3HC4 type (RING finger)
HMM *CPICFCtTFQldyPWPfdePmMlPCgHsFCypCIrrW..CPmC*
C++C + F++P++++CGH+FC+ C +++ CP+
Query 55 CQLC-----CSV---FKDPVITTCGHTFCRCALKSEKCPVD 88

DKFZphutel_20b19

group: metabolism

DKFZphutel_20b19 encodes a novel 486 amino acid protein with similarity to bacterial sarcosine oxidases (EC 1.5.3.1.)

The novel protein seems to be a novel enzyme with sarcosine oxidase activity.

The new protein can find application in modulation of sarcosine metabolism and as a new enzyme for biotechnologic production processes.

similarity to sarcosine oxidases

membrane regions: 1

Summary DKFZphutel_20b19 encodes a novel 486 amino acid protein, with similarity to sarcosine oxidases.

similarity to sarcosine oxidases

complete cDNA?, complete cds potential start at Bp 48, EST hits.

Sequenced by AGOWA

Locus: unknown

Insert length: 1967 bp

Poly A stretch at pos. 1950, no polyadenylation signal found

```
1 AGCGAGGCAG CAGTGCAGCT TTCAGAGGGT CCGGGCTCAG AGGGGTTATG
51 ATTCCGAGGG TTCTGCCGCA CGGCATGGGC CGGGGCTCT TGACCCGGAG
101 GCCAGGCACG CGCAGAGGAG GCTTTTCTCT GGACTGGGAT GGAAAGGTGT
151 CTGAGATTAA GAAGAAGATC AAGTCGATCC TGCCCTGGAAG GTCTGTGAT
201 CTACTGCAAG ACACCAAGCCA CCTGCCTCCC GAGCACTCGG ATGTGGTGAT
251 CGTGGGAGGT GGGGTGCTTG CTTTGTCTGT GGCTATTGG CTGAAGAAGC
301 TGGAGAGCAG ACGAGGTGCT ATTCGAGTGC TASTGGTGA ACGGGACCAC
351 ACCTATTCAC AGGCTCTCAC TGGGCTCTCA GTAGTGGA TTGTGACGA
401 GTTCTCATTG CCTGAGAACA TCCAGCTCTC CCTCTTTTCA GGCAGCTTTC
451 TACCGAATAC CAATGAGTAC CTGGCCCTAG TCGATGCTCC TCCCTGGAC
501 CTCGGTTTCA ACCCTCTGGG CTACCTCTTG CTGGCTTCAG AAAAGGATGC
551 TGCAGCCATG GAGAGCAACG TGAAAGTGCA GAGGCAGGAG GGAGCCAAAG
601 TTTCTCTGAT GTCTCCTGAT CAGCTTCGGA ACAAGTTTCC CTGGATAAAC
651 ACAGAGGGAG TGGCTTTGGC GTCTTATGGG ATGGAGGACG AAGGTTGGTT
701 TGACCCCTGG TGTCTGCTCC AGGGGCTTCG GCGAAAGTCC CAGTCTTTGG
751 GAGTCCTTTT CTGCCAGGGA GAGGTGACAC GTTTTGTCTC TTCATCTCAA
801 CGCATGTTGA CCACAGATGA CAAAGCGGTG GTCTTGAAA GGATCCATGA
851 AGTCATGTG AAGATGACCC GCAGCCTCGA GTACCACTCT GTGGATGCG
901 CCATTGTGAT CAACGCAGCC GGAGCCTGGT CTGCGCAAT CGCAGCACTG
951 GCTGGTGTG GAGAGGGGCC GCCTGGCACC CTGCAGGGCA CCAAGCTACC
1001 TGTGGAGCCG AGGAAAGGT ATGTGTATGT GTGGCACTGC CCCCAGGGAC
1051 CAGGCCTAGA GACTCCGCTT GTTGACAGA CCAGTGAGGC CTATTTTCGC
1101 CGGGAAGGAT TAGGTAGCAA CTACCTAGGT GGTCTAGGCC CCACTGAGCA
1151 GGAAGAACC GACCCGGCGA ACCTGGAAGT GGACCATGAT TTCTTCCAGG
1201 ACAAGGTGTG GCCCCATTGT GCCCTGAGGG TCCAGCTTT TGAGACTCTG
1251 AAGGTTTACA GCGCCTGGGC CGGCTATTAC GACTACAACA CCTTTGACCA
1301 GAATGGCTG GTGGGCCCC ACCCGCTAGT TGTCACATG TACTTTGCTA
1351 CTGGCTTCAG TGGTACGGG CTCCAGCAGG CCCCTGGCAT TGGGCGAGCT
1401 GTAGCAGAGA TGGTACTGAA GGGCAGGTT CAGACCATG ACCTGAGCCC
1451 CTTCTCTTT ACCCGCTTTT ACTTGGGAGA GAAGATCCAG GAGAACAACA
1501 TCATCTGAGC ATGTGTGCTC TGCATGGCT CCAGTGGCTT GCATCCTGGC
1551 TGTGTTTACA GCCTTGTGTT GTGCTTCCAT CTTCCCCAGT ACTGTGCCAG
1601 GCCTTCTCCC CCTCCCCAGT GTCTCTCCT CTCAGGCAGG CCATTGCACC
1651 CATATGGCTG GCGAGGCACA GGCAGTGAGG CCGAGGCCAA TAGCGAGTGA
1701 TGAGCGGGAT CCTAGGACTG ATCTGTAGCC CATGCTGATG TCACCCACCA
1751 GGGCAATCCA TCTGGAGGCC TGAGCACCTT GGCCTCAGGAC TGGCTTCATC
1801 CTGGCACTGA CCAGGAAGCA CTGGCTTGA CCCCTTACG AGACAGAGCC
1851 CAGGCACTGG AGCACTCTGG GCGAGCCTGG CTCAGGTTTA TTGATTTTCG
1901 TCTGTTTACC CTATCCATTA ATCAATACAT GTAATTAATC CTTTCCCTCC
1951 AAAAAAAAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 48 bp to 1505 bp; peptide length: 486
Category: similarity to known protein

```
1 MIRRVLPHGM GRGLLTRPG TRGGFSLDW DGKVEIKKK IKSILPGRSC
51 DLLQDTSHLP PEHSDVIVG GGVGLGSVAY WLKLESRRG AIRVLVVERD
101 HTYSQASTGL SVGGICQQFS LPENIQLSLF SASFLRNINE YLAVVDAPPL
151 DLRFPNSGYL LLASEKDAAA MESNVKVRQ EGAKVSLMSP DQLRNKFPMI
201 NTEGVALASY GMEDEGWFDW WCLLQGLRRK VQSLGVLFQ GEVTRFVSS
251 QRMLTTDDKA VVLKRIHEVH VKMORSLEYQ PVECAIVINA AGAWSAQIAA
301 LAGVGEPPG TLQGTLPVE PRKRIVVVMH CPOGPGLET PLVADTS-G
351 RREGLSNYL GGRSPTEQEE DPANLEVDH DFFQDKVNP LALRVPAFET
401 LKVSQAWAGY YDNTFDQNG VVGPHPLVM MYFATGFSGH GLQAPGIGR
451 AVAEMVLKGR FQTIDLSFPL FTRFYLGEKI QENNII
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1_20b19, frame 3

TREMBL:CEM04B2.4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2,
N = 1, Score = 801, P = 9.2e-80

PIR:B71184 probable sarcosine oxidase - Pyrococcus horikoshii, N = 2,
Score = 194, P = 2e-26

PIR:B69284 sarcosine oxidase, subunit beta (soxB) homolog -
Archaeoglobus fulgidus, N = 3, Score = 189, P = 8.2e-22

TREMBL:AF042732.1 gene: "Bb"; product: "unknown protein"; Anopheles
gambiae (Bb) gene, partial cds; and TU37B2 (TU37B2) and diphenol
oxidase-A2 (Dox-A2) genes, complete cds., N = 1, Score = 386, P =
8.7e-36

PIR:F71008 probable sarcosine oxidase - Pyrococcus horikoshii, N = 2,
Score = 200, P = 4e-25

>TREMBL:CEM04B2.4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2
Length = 527

HSPs:

Score = 801 (120.2 bits), Expect = 9.2e-80, P = 9.2e-80
Identities = 171/433 (39%), Positives = 260/433 (60%)

```
Query: 61 PEHSDVIVGGGVGLGSVAYWLKLESRRGAIKRLVVRDHTYSQASTGLSVGGICQQFS 120
P +++VI+GGG+ G S A+WLK+ R +++++VVE + +++ST LS GGI QQFS
Sbjct: 91 PYRAEIVVGGGLSGSSTAFWLKE-RFADEDFKVVVVENNNDVFTKSTMLSTGGITQQFS 149

Query: 121 LPENIQLSLFSASFRLRNINEYLAVVDAPPLDLRFNPSPGYLLA-SEKDAAMESNVKVR 179
+PE + +SLF+ FLR+ E+L ++D+ D+ F P+GYL LA +++ M S KVQ
Sbjct: 150 IPEFVDMSLFTTEFLRHAGEHLRILOSEQPDINFPTGYLRRLAKTDEEVMMRSAAWKVQI 209

Query: 180 QEGAKVSLMSPDQLRNKFPWINTGVALASYGMEDEGWFDWPCLLQGLRRKQSLGVLF 239
+ GAKV L+S D+L ++P++N + V LAS G+E+EG D W LL +R K +LGV +
Sbjct: 210 ERGAKVQLSKDELTKRYPIYMNVDVLLASLGVENEGTIDTWQLLSAIREKNITLGVQYV 269

Query: 240 QGEVTRFVSSSRH-----LTTDDKAVVLKRIHEVHVMDRS-LEYQPECAIVI 288
+GEV F R T D+ + +RI V V+ + +P+ +++
Sbjct: 270 KGEVEGFQFERHRASSEVHAFGDDADENKLRQIRISGVLRPQMNDASARPIRAHLIV 329

Query: 289 NAAGAWSAQIAALAGVGEPPGTQLQGTLPVEPRKRIVVVMHCPQPGLETPLVADTS-G 347
NAAG W+ Q+A +AG+G+G G L +P++PRKR V+V P P + P + D S G
Sbjct: 330 NAAGPWAGQVAKMAGIGKGT-GLL-AVVPVPIQPRKRDVFVIFAPDVPS-DLPFIIDPSTG 386

Query: 348 AYFRREGLSNYLGGSPTEQEE--DPANLEVDHDFQDKVNPMLALRVPAFETLKVQS 405
+ R+ G +L GR+P+++E+ D +NL+VD+D F K+WE L RVP F+T KV+S
Sbjct: 387 VFGRQTDGQTFVLVGRTPSKEEDAKRHSNLDVYDFYQKIWPVLVDRVPGFTAKVKS 446
```

Query: 406 AWAGYYDYNTFDQNGVVGPHPLVNNMYFATGFSGHGLQQAAGIGRAVAEMVLKGRFOTID 465
 AW+GY D NTFD V+G HPL N++ GF G+ + RA AE + G + ++
 Sbjct: 447 AWSGYQDINTFDDAPVIGEHPLYTNLHNMCGFGERGMHMAAARAYAEIRIFDGAYINVN 506

Query: 466 LSPFLFTRFYLGEKIQE 482
 L F R + I E
 Sbjct: 507 LRKFDMMRRIVKMDPITE 523

Pedant information for DKFZphute1_20b19, frame 3

Report for DKFZphute1_20b19.3

[LENGTH] 486
 [MW] 53811.85
 [PI] 7.66
 [HOMOL] TREMBL:CEM04B2_4 gene: "M04B2.4"; Caenorhabditis elegans cosmid M04B2 1e-78

[FUNCAT] c energy conversion [H. influenzae, HI0499] 8e-05
 [BLOCKS] BL00677A D-amino acid oxidases proteins
 [BLOCKS] BL00623A GMC oxidoreductases proteins
 [BLOCKS] BL01304A
 [EC] 1.5.99.2 Dimethylglycine dehydrogenase 2e-07
 [PIRKW] flavoprotein 2e-07
 [PIRKW] oxidoreductase 2e-07
 [PROSITE] MYRISTYL 12
 [PROSITE] CK2_PHOSPHO_SITE 5
 [PROSITE] GLYCOSAMINOGLYCAN 1
 [PROSITE] PKC_PHOSPHO_SITE 6
 [KW] TRANSMEMBRANE 1
 [KW] LOW_COMPLEXITY 7.00 %

SEQ MIRRVLPHGMGRGLLTRPGTRGGFSLDWDGKVEIKKKIKSILPGRSCDLLQDTSHP
 SEGXXXXXXXXXXXXXXXXX.....XXXXXXXXX.....
 PRD cccceccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccc
 MEM

SEQ PEHSDVIVGGVGLSVAYMLKKLESRRGAIRVLVVERDHTYSQASTGLSVGGICQQFS
 SEGXXXXXXXXXXXXX.....
 PRD cccceccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccc
 MEMMMMMMMMMMMMMMMMM.....

SEQ LPENIQLSLSFASFLRNINEYLAVVDAAPLDLRFNPSGYLLASEKDAAMESNVKVRQ
 SEG
 PRD cchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccc
 MEM

SEQ EGAKVSLMSPDQLRNKFPWINTGVALASYGMEDEGWDFPMCLLQGLRRKVQSLGVLPQ
 SEG
 PRD cccceccccchhhhhccccccccccccccccccccccccccccchhhhhhhhhhhheeeec
 MEM

SEQ GEVTRFVSSQRMLTTDDKAVVLKRIEHHVHKMDRSLEYQPECAIVINAAGAWSAQIAA
 SEG
 PRD cccccccccccccccccchhhhhhhhhhecccccccccccccccccccccccccccc
 MEM

SEQ LAGVGEGPPGTQGTLPVEPRKRYVYVMHCPOGPLETPLVADTSGAYFRREGLSNYL
 SEG
 PRD hhcc
 MEM

SEQ GGRSPTEQEEPDPANLEVDHDFQDKVMPHLALRVPAFETLVQSAWAGYYDYNTFDQNG
 SEG
 PRD cccccccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccc
 MEM

SEQ VVGPHPLVNNMYFATGFSGHGLQQAAGIGRAVAEMVLKGRFOTIDLSPFLFTRFYLGEKI
 SEG
 PRD cccccccccccccccccchhhhhhhhhhhhhcccccccccccccccccccccccc
 MEM

SEQ QENNII
 SEG
 PRD ccccc
 MEM

Prosites for DKF2phutcl_20b19.3

PS00002	438->442	GLYCOSAMINOGLYCAN	PDOC00002
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	87->90	PKC_PHOSPHO_SITE	PDOC00005
PS00005	164->167	PKC_PHOSPHO_SITE	PDOC00005
PS00005	250->253	PKC_PHOSPHO_SITE	PDOC00005
PS00005	400->403	PKC_PHOSPHO_SITE	PDOC00005
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00006	364->368	CK2_PHOSPHO_SITE	PDOC00006
PS00006	366->370	CK2_PHOSPHO_SITE	PDOC00006
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	75->81	MYRISTYL	PDOC00008
PS00008	109->115	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	204->210	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	292->298	MYRISTYL	PDOC00008
PS00008	310->316	MYRISTYL	PDOC00008
PS00008	354->360	MYRISTYL	PDOC00008
PS00008	447->453	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phutcl_20b19.3)

DKF2phut1_20g21

group: signal transduction

DKF2phut1_20g21 encodes a novel 861 amino acid protein with partial similarity to human ras inhibitor and other ras inhibitor proteins.

Ras is a signal transducing molecule involved in the receptor tyrosine kinase/RAS/Map kinase signalling cascade. Ras proteins bind GDP/GTP and show intrinsic GTPase activity. Mutations in ras, which change aa 12, 13 or 61 activate the potential of ras to transform cultured cells and are implicated in a variety of human tumours. The novel protein seems to be a new ras inhibitor protein.

The new protein can find application in modulating/blocking ras dependent signal transduction pathways.

Ras inhibitor

additional 1188 Bp at 5' and 1107 at 3' end in comparison to I22483

Sequenced by AGOWA

Locus: unknown

Insert length: 4137 bp
Poly A stretch at pos. 4116, no polyadenylation signal found

```
1 GGGAGAACTG AAACAGGAGA TGGTGGGAC AGATGTCAAC CTGGAAAAATG
51 GCCTGGAACC CGCTGAAACC CACAGCATGG TAAGACACAA GGATGGTGGC
101 TATTCGAGG AAGAGGACGT GAAGACCTGT GCCCGGAGCT CAGGCTATGA
151 CAGCCTCTCC AACAGGCTCA GCATCTTGA CCGGCTCCTC CACACCCACC
201 CCATATGGCT GCAGCTGAGT CTGAGTGAGG AGGAGGCAGC AGAGGTCCTG
251 CAGGCCCAGC CTCGGGGGAT CTTCTGGTT CATAAATCTA CCAAGATGCA
301 GAAGAAAGTC CTCCTCCCTCC GCCTGCCCTG TGAATTGGG GCCCCACTCA
351 AGGAATTTGC CATAAGGAA AGCACAATCA CTTTTCCTT GGAAGGCTCA
401 GGAATCAGTT TCGCAGATT ATTCCGGCTC ATTGCTTTCT ACTGCATCAG
451 CAGGATGTTT CTACCATTTA CTTGAAGTT GCCTTATGCC ATTTCAACAG
501 CCAAGTCGGA GGCTCAGCTT GAAGAAGTGG CCCAGATGGG ACTAAATTTT
551 TGGAGCTCCC CAGCTGACAG CAACCCCCCG AACCTTCCAC CTCCCATAG
601 GCCTCTTTCC TCCGACGGTG TCTGTCTGCT CTCCTTGGCT CAGCTCTGCC
651 TTATAAATGG AGTGCAATCT ATCAAAACCA GGACGCCCTC AGAGCTGGAG
701 TGCAGCCAGA CCAACGGGGC CCGTGTGTTT ATTAATCCCC TTTTCTTGAA
751 AGTGACACAG CAGGACCTCA GTGGAGGCTT GAAACGGCCG AGCACAAGGA
801 CTCCTCAACG GAATGGCAGG GAGCGGACTC GTTCCCCCCC ACCCAGGCC
851 CCGCCACCCG CTATTATAG TCTCCACACA AGCCCTGGGC TGCCACGGAC
901 TGAACCCAG ACCGCAATGC CAGAAACAGT CAACATAAC AAACATGGGA
951 ACGTAGCTCT CCTTGAACG AAACCAACTC CCATCCCTCC ACCCGGGCTG
1001 AAGAAGCAGG CTTCTTTTCT GGAAGCAGAG GCGGCTGCAA AGACCTTGAG
1051 GCGCGGCCGG CCGGGGCGAG GCCCGAGCTC GAGCTGGGAC ACAGCTGGCA
1101 GCCCAGGTGG GCGCCCGCCT GAGGCGGCCG CCGGGGATTC CACAAGGGCC
1151 CCGCCGCCCA GCTCTGAATC ACGGCCCGCG TGCCATGGAG GCCGGCAGCG
1201 GCTGAGCGAC ATGAGCATTT CTACTTCCCT CTCGACTCG CTGGAGTTCT
1251 ACCGGAGCAT GCCTCTGTTT GGCTACGAGG CGGACACCAA CAGCAGCCTG
1301 GAGGACTACG AGGGGGAAG TGACCAAGAG ACCATGGGCG CCCCATCAAA
1351 GTCCAAAGAG AAAGAGGACA GCTCTTCTGT CCTGCCAAG CTCTCAACT
1401 CCCAGCTGCA GAAGGTGAGC GGGGTGTTCA GCTCTTCAT GACCCCGGAG
1451 AAGCGGATGG TCCCGAGGAT CGCCGAGCTT TCCCGGACAA AATGCACCTA
1501 CTTGGGTGCT TTAGTGCAGG ACTACGTGAG CTTCTCTGAG GAGAACAAGG
1551 AGTGCCACGT GTCCAGCACC GACATGCTGC AGACCATCCG GCAGTTTCATG
1601 ACCCAGGTCA AGAACTATT GTCTCAGAGC TCGGAGCTGG ACCCCCCCAT
1651 CGAGTCGCTG ATCCCTGAAG ACCAAATAGA TGTGCTGCTG GAAAAAGCCA
1701 TGCACAAAGT CATCTTGAAG CCCCACAGG GGCATGTGGA GGCCATGCTG
1751 AAGGACTTTC ACATGGCCGA TGGCTCATGG AAGCAACTCA AGGAGAACCCT
1801 GCAGCTTGTG CCGCAGAGGA ATCCGAGGA GCTGGGGGTC TTGCCCCGA
1851 CCCCCTGATT TGTGGATGTG GAGAAATCA AAGTCAACTT CATGACCATG
1901 CAGAAGATGT ATTCCGCCGA AAAGAAGTTC ATGCTGCTGC TCGCGGCTCTG
1951 CAGCTCATTT TACACGGTCA TGGAGAACAA CTCAGGGAGG ATGTATGGCG
2001 CTGATGACTT CTTGCCAGTC CTGACCTATG TCATAGGCCA GTGTGACATG
2051 CTTGAATTGG AACTGAAAT CGAGTACATG ATGGAGCTCC TAGACCCATC
2101 GCTGTATACAT GGAGAAGGAG GCTATTACTT GACAAGCGCA TATGGAGCAC
2151 TTTCTCTGAT AAAGAATTTT CAAGAAGAAC AAGCAGCGCG ACTGCTCAGC
2201 TCAGAAACCA GAGACACCTT GAGGCACTGG CACAAACGGA GAACCACTAA
2251 CCGGACCATC CCCTCTGTGG ACGACTTCCA GAATTACCTC CGAGTTGCAT
2301 TTCAGAGGCT CAACAGTGGT TGCACAGGAA AGACCTCTCT TGTAGACCTT
2351 TACATCACC ACTGAGATCT GTGTCAGATC TGCCTGAGA AGTTCAAGGT
2401 GCGGGACCTT GAGGAGTACA GCCTTTTCT CTTCTTTGAC GAGCATGGC
2451 AGCAGCTGGC AGAGGACACT TACCCTCAAA AATCAAGGC GGAGCTGCAC
```

```
2501 AGCCGACCAC AGCCCCACAT CTTCACATTT GTCTACAAAC GCATCAAGAA
2551 CGATCCTTAT GGCATCATTT TCCAGAACGG GGAAGAAGAC CTCACCACCT
2601 CCTAGAAGAC AGCGGGGACT TCCAGTGGT GCATCCAAAG GGGAGCTGGA
2651 AGCCTTGCTT TCCGCTTCT ACATGCTTGA GCTTGAAAG CAGTCACCTC
2701 CTCGGGGACC CCTCAGTGTA GTGACTAAGC CATCCACAG CCAACTCGGC
2751 CAGGGCAAC TTTAGCCACG CAGGTAGCT GAGGTTTTGT AAACAGTAGG
2801 ATTCTCTTTT GGCAATGGAG AATTGCATCT GATGGTCAA GTCTCCTGAG
2851 ATTGTTTGT ACCTACCCCC AGTCAGTTT TAGGTTGGCT TACAGGTATG
2901 TATATGTGCA GAAGAACAC TTAAGATACA AGTTCTTTTG AATTCAACAG
2951 CAGATGCTTG CGATGCAGTG CGTCAGGTGA TTCTCACTCC TGTGGATGGC
3001 TTCATCCCTG CCTTCCTTCC TTTCCTTTTC CTTTTTTTTT TTTTTTTTTT
3051 TTTTACAAA GAGCCTTCAT GTTTTATAT ATTTATAGA AATTTTATA
3101 GCAGTTGCAG GTAAACTGTC AGGATTGGTT TTAATATATT TTGTAACTT
3151 TAAATATTC TAATAATTG CATGTGATT TAACATTATA TATTCAAAA
3201 TAAATCTCTT GCTGGATTG AGAGTATTGC ATTTTAAAG TCTCTCTCT
3251 GTAATGGAT GTTTTGGCAA CTTTGTGGGG AGAGACTGCT GGATTTCTTA
3301 AAGCAACGTA TTCTGCACAC TGGCCACAGA ATGCTTTGG AAATCGGATG
3351 TACTGTCTCT TTGTTGACGT TTAGTGGTGT TTTGCTGTTT TGTTTTTTAA
3401 ACAAATGATG CTGAGATAA GGAGAGAAAT GAATGTAGAG AGAGGTAGAG
3451 AGAGAAATAT GAACCTAAC AAAGGACTGA GGAGTGCAGT CTGCTGGTTC
3501 AGGCTCTTCA AAGATGTAG AAAAGAGAT AGAAGGAACC ACCTATGCTT
3551 AAAATCTGT AAATATGCAG TGAGGTTTGG CAAATCTAT TCCATGTGTG
3601 ATTGCTTGT AGAAACAATT TGAAGCGC CTGAGGAAA ATAAAATCA
3651 AGAAGACAC TTTCCTCCT TTTCATACA AATTAAACT TAACAGCATC
3701 AATATTGGG GACCAGAAAC CAGTAATGT ATAATGTGGC TTTTGTGAG
3751 TTAATAAGA TGCTATATA TGGAGAAGAA TTTGAAATG CACAAAAAAA
3801 TCAATCTACA TTATCAGAAC CTGAGTGAA ATTAACATTA TGTAAATAA
3851 AACCAAGTTG CAGGTGCACA AACTATGAGG GTCTGTATC CAGTAACAC
3901 AGGTAGTTAC AAAAACATGT TATTGTACTG TGAAGATG CATAGTCATC
3951 TCATTGTTT GGCTTTGTAC CTTGTACCTT TTTTAGCCTT GGCTTTTGT
4001 GAATCAGAAC CCTCAGCACA TACTGTGTTG TACTTTTGT AATGATTTT
4051 TAAATGGAAT TTGACATA ATACATTGTA ATACTGTATG ATAATCATG
4101 GTGAAATAA TTTTGAAT AAAAATAA AAAAAA
```

BLAST Results

Entry 122483 from database EMBL:
Sequence 15 from patent US 5527896.
Length = 1829
Plus Strand HSPs:
Score = 9097 (1364.9 bits), Expect = 0.0, P = 0.0
Identities = 1821/1823 (99%), Positives = 1821/1823 (99%),

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2602 bp; peptide length: 861
Category: known protein
Classification: Cell signaling/communication

```
1 MVRTDVNLEN GLEPAETHSM VRHKDGGYSE EEDVKTCARD SGYDLSNRL
51 SILDRLHMT PIWLQSLSE EEAEEVLQAO PPGIFLVHKS TKMKQKVLSL
101 RLPCEPGAPL KEFAIKESTY TFSLEGGIS FADLFRLIAF YCISRDVLPF
151 TLKLPYATST AKSEAQLEEL AQMGLEWSS PADSKEPMLP PPHRPLSDG
201 VCPASLRQLC LINGVHSIKT RTPSELECSQ TNGALCFINP LFLKVHSDQL
251 SGGKLRPSTR TPNANGTERT RSPPPRPPPP AINSLHTSPR LARTETQSM
301 PETVNMHKG NVALPGTKPT PIPPPRLKKQ ASFLEAEGGA KTLGGRRPGA
351 GPELELGTAG SPGGAPPEAA PGDCTRAPPP SSESRRPCHG GRQLSDMSI
401 STSSSDSLEF DRSMPLFGYE ADTNSSLEDY EGESDOETMA PPIKSKKRS
451 SSFVLPKLVK SQLQKVGVSF SSFMTPEKRM VRRIAELSRO KCTYFGCLVQ
501 DYVSFLQENK ECHVSSTDML QTIROFMTQV KNYLSQSSEL DPPIESLIPE
551 DQIDVLEKA MHKCILKPLK GHVEAHLKDF HMADGSWKQL KENLQLVQR
601 NPQELGVFAP TPDVDEKFI KVFMTQKRM YSPEKVMALL LRVCKLIYTV
651 MENNSGRNYG ADDFLPVLTY VIAQDMLEL DTEFLWMEL LDPSLLHGG
701 GYVLSAYCA LSLIKNFQEE QAARLLSSET RDTIRQWHR RTNRTIPSV
751 DDFQNYLRVA FOEVNSGCTG KILLVRPYIT TEDVQICAE KFKVGDPEEY
801 SLFLVDETW QQLAEDTYPQ KIKAEHSRP QPHIFRVYK RIKNDPYGII
851 FQNGEEDLTT S
```


BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutcl_20g21, frame 2

TREMBL:RNU80076_1 product: "RIN1"; Rattus norvegicus RIN1 mRNA,
complete cds., N = 3, Score = 606, P = 6.8e-97

PIR:A38637 Ras interactor RIN1 - human, N = 3, Score = 587, P = 1.9e-92

TREMBL:HSRASINL_1 product: "ras inhibitor"; Human ras inhibitor mRNA,
3' end., N = 2, Score = 592, P = 9.8e-61SWISSPROT:RIN1_HUMAN RAS INTERACTION/INTERFERENCE PROTEIN 1 (RAS
INHIBITOR JC99) (FRAGMENT) ., N = 2, Score = 587, P = 4.1e-60PIR:B38637 Ras inhibitor (clone JC265) - human (fragment), N = 1, Score
= 2446, P = 4.6e-254>PIR:B38637 Ras inhibitor (clone JC265) - human (fragment)
Length = 471

HSPs:

Score = 2446 (367.0 bits), Expect = 4.6e-254, P = 4.6e-254
Identities = 471/471 (100%), Positives = 471/471 (100%)Query: 391 GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETHAPPIKSKKKRS 450
GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETHAPPIKSKKKRS
Sbjct: 1 GRQLSDMSISTSSSDSLEFDRSMPLFGYEADTNSSLEDYEGESDQETHAPPIKSKKKRS 60Query: 451 SSFVLPLKLVKSQQLQKVGVSFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 510
SSFVLPLKLVKSQQLQKVGVSFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK
Sbjct: 61 SSFVLPLKLVKSQQLQKVGVSFSSFMTPEKRMVRRIAELSRDKCTYFGCLVQDYVSFLQENK 120Query: 511 ECHVSSTDMQLQTIHQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLKAMHKCILKPLK 570
ECHVSSTDMQLQTIHQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLKAMHKCILKPLK
Sbjct: 121 ECHVSSTDMQLQTIHQFMTQVKNYLSQSSELDPPIESLIPEDQIDVVLKAMHKCILKPLK 180Query: 571 GHVEAMLKDFHMDGSKWQKLENLQVQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 630
GHVEAMLKDFHMDGSKWQKLENLQVQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM
Sbjct: 181 GHVEAMLKDFHMDGSKWQKLENLQVQRNPQELGVFAPTPDFVDVEKIKVKFMTMQKM 240Query: 631 YSPEKKVMILLRVCKLIYTMENNSGRMGADDFLPVLTYYIAQCDMLELDEIEYMME 690
YSPEKKVMILLRVCKLIYTMENNSGRMGADDFLPVLTYYIAQCDMLELDEIEYMME
Sbjct: 241 YSPEKKVMILLRVCKLIYTMENNSGRMGADDFLPVLTYYIAQCDMLELDEIEYMME 300Query: 691 LDPSLLHGEHGYLTSAAGALSILKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV 750
LDPSLLHGEHGYLTSAAGALSILKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV
Sbjct: 301 LDPSLLHGEHGYLTSAAGALSILKNFQEEQAARLLSSETRDTLRQWHKRRRTNRTIPSV 360Query: 751 DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLVDET 810
DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLVDET
Sbjct: 361 DDFQNYLRVAFQEVNSGCTGKTLVLRPYITTEDVCQICAEKFKVGDPEEYSLFLVDET 420Query: 811 QQLAEDTYPQIKAEHLSRPQPHIFHVFYKRIKNDPYGIIFQNGEEDLTTS 861
QQLAEDTYPQIKAEHLSRPQPHIFHVFYKRIKNDPYGIIFQNGEEDLTTS
Sbjct: 421 QQLAEDTYPQIKAEHLSRPQPHIFHVFYKRIKNDPYGIIFQNGEEDLTTS 471

Pedant information for DKFZphutcl_20g21, frame 2

Report for DKFZphutcl_20g21.2

[LENGTH]	861
[MW]	96380.26
[PI]	6.15
[HOMOL]	PIR:B38637 Ras inhibitor (clone JC265) - human (fragment) 0.0
[FUNCAT]	08.13 vacuolar transport [S. cerevisiae, YML097c] 3e-10
[FUNCAT]	06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]
[FUNCAT]	3e-10
[FUNCAT]	30.03 organization of cytoplasm [S. cerevisiae, YML097c] 3e-10
[FUNCAT]	08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]
[FUNCAT]	3e-10
[PIRKW]	alternative splicing 3e-59
[SUPFAM]	Ras interactor RIN1 3e-59

[illegible]

(No Prosite data available for DKF2phut1_20g21.2)

(No Pfam data available for DKF2phut1_20g21.2)

DKFZphut1_20h13

group: intracellular transport and trafficking

DKFZphut1_20h13 encodes a novel 955 amino acid protein with similarity to alpha-adaptins.

Adaptins are components of the adaptor complexes which link clathrin to receptors in coated vesicles. The alpha-adaptins, which are found exclusively in endocytic coated vesicles, separate into two bands on SDS gels, designated A and C. The novel protein is very similar to both alpha adaptin A and C. The novel protein is a new human alpha-adaptin.

The new protein can find application in modulating endocytosis and vesicle trafficking in cells.

strong similarity to alpha-adaptins

complete cDNA, complete cds start at Bp 78, EST hits

Sequenced by AGOWA

Locus: unknown

Insert length: 3352 bp

Poly A stretch at pos. 3297, polyadenylation signal at pos. 3279

```
1 GGGCCCGGTC CCCGCTTGCC AGCCCCCGCT GCTCTGTGCC CTGTCCGGCC
51 AGGCCCTGGAG CCGACACCAC CGCCATCATG CCGGCCCTGT CCAAGGGCGA
101 TGGGATGCGG GGGCTCGCGG TGTTTCATCT CGACATCCGG AACTGTAAAG
151 GCAAAAGAGG GGAATATTAG AGAATCAACA AGGAACCTGG CAACATCCGC
201 TCCAAAGTTCA AAGGAGACAA AGCCTTGGAT GGCTACAGTA AGAAAAATA
251 TGTGTGTAAA CTGCTTTTCA TCTTCTGCT TGGCCATGAC ATTGACTTTG
301 GGCACATGGA GGCTGTGAAT CTGTGTAGTT CCAATAAATA CACAGAGAAG
351 CAAATAGGTT ACCTGTTTCA TTCTGTGCTG GTGAACCTGA ACTCGGAGCT
401 GATCCGCGCT ATCAACAACG CCATCAAGAA TGACCTGGCC AGCCGCAACC
451 CCACCTTTCAT GTGCTGGGCC CTGCACTGCA TCGCCAACGT GGGCAGCCGG
501 GAGATGGCGG AGGCTTTTGC CGCTGACATC CCCCSCATCC TGGTGGCGGG
551 GGACAGCATG GACAGTTTCA AGCAGAGTGC GGGCCTGTGC CTCTTCGAC
601 TGTACAAGGC CTGCGCTGAC CTGGTGCCCA TGGCGGAGTG GACGGCGCGT
651 GTGGTACACC TGCTCAATAG CCAGCACATG GGTGTGGTCA CGGCCGCGGT
701 CAGGCTCATC ACCTGTCTCT GCAAGAAGAA CCCAGATGAC TTCAAGACGT
751 GCGTCTCTCT GGCTGTGTGC CGCTGAGGCC GGATCGTCTC CTCTGCCTCC
801 ACCGACCTCC AGGACTACAC CTACTACTTC GTCCACGAC CCGGCTCTC
851 GGTGAAGCTC CTGCGGCTGC TGCAGTGCTA CCGGCTTCCA GAGGATGGGG
901 CTGTGAAGGG CCGGCTGGTG GAATGTCTGG AGACTGTGCT CAAACAAGGC
951 CAGGAGCCCC CCAATCCAA GAAGGTGCAG CATTCCAAG CCAAGAACGC
1001 CATCTCTTTC GAGACATCA GCCTCATCAT CCACTATGAC AGTGAGCCCA
1051 ACCTCCTGGT TCGGCGCTGC AACCAGCTGG GCCAGTTCTT GCAGCACCGG
1101 GAGACCAACC TCGGCTACCT GGGCCTGGAG AGCATGTGCA CGCTGGCCAG
1151 CTCGAGTTTC TCCCATGAAG CCGTCAAGAC GCACATTGAC ACCGTATCA
1201 ATGCCCTCAA GACGAGCGGG GACGTACGGG TCGGCGAGCG GCGGCTGAC
1251 CTCTCTACGG CCAATGTGTA CCGGAGCAAT GCCAAGCAGA TGGTGTGGA
1301 GATGCTGCGG TACTTGAGA CCGCAGACTA CGGCATCCGC GAGGAGTCC
1351 TCCTGAAGGT GGCCATCTCT GCGGAGAGT AGCGCTGTGA CTACAGCTGG
1401 TACGTGACA CCACTCTCAA CCTCATCCGC ATTGCGGGCG ACTACGTGAG
1451 TGAGGAGGTG TGGTACCGTG TGCTACAGAT CGTCACCAAC CGTGATGAGC
1501 TCCAGGGCTA TCGCGCCAAG ACCGTCTTTG AGGCGCTCCA GGCCCTTCCC
1551 TGTCACGAGA ACATGGTGA GGTGGCGGGC TACATCTTGG GGGAGTTTGG
1601 GAACCTGATT GCTGGGGACC CCGGCTCCAG CCCCCAGTG CAGTTCTCCC
1651 TGCTCCACTC CAAGTTCCAT CTGTGACGGG TGGCAGGGGG GGGCTGCTG
1701 CTGTCCACCT ACATCAAGTT CATCAACCTC TTCCCGAGA CCAAGGCCAC
1751 CATCCAGGGC GTCTGCGGG CCGGCTCCCA GCTGCGCAAT GCTGACGTGG
1801 AGCTGCAGCA GCGAGCCGTG GAGTACCTCA CCTCAGCTC AGTGGCCAGC
1851 ACCGACGTCC TGGCCACGGT GCTGGAGGAG ATGCCGCCCT TCCCGGAGCG
1901 CGAGTCGTCC ATCTCGGCCA AGCTGAAACG CAAGAAGGGG CCAGGGGCCG
1951 CGAGCGCCCT GGACGATGGC CGGAGGGACC CCAGCAGCAA CGACATCAAC
2001 GGGGGCATGG AGCCCAACCC CAGCACTGTG TCGACGCCCT CGCCCTCCGC
2051 CGACCTCTCT GGGCTGCGGG CAGCCCTTCC CCGGCGAGCA CCCCCGCTT
2101 CTGCGGAGGC AGGGAACCTT CTGGTGAGC TCTTCGATGG CCGGCGGCC
2151 CAGCCAGGCC TGGGGCCAC CCGGAGGAG GCTTCTCTCA GCGCAGGTCC
2201 TGAGGACATC GGCCCTCCCA TTCCGGAAGC CGATGAGTTG CTGAATAAGT
2251 TTGTGTGTAA GAACAACGGG GTCTGTGTC AGAACCAGCT GCTGCAGATC
2301 GGAGTCAAGT CAGAGTTCCG ACAGAACCTG GCGCGCATGT ATCTCTCTTA
2351 TGGCAACAAG ACCTCGGTGC AGTTCAGAA TTCTCACCC ACTGTGTTT
2401 ACCCGGAGAA CCTCCAGACT CAGCTGGCTG TGCAGACCAA GCGCGTGGCG
2451 GCGCAGGTGG ACGGCGGGCG GCAGGTGCAG CAGGTGCTCA ATATCGAGTG
2501 CCGCGGGGAC TTCTGACGC CCGGCTGCT GTCCGTGCGC TTCCGTAACG
2551 GTGGCGGCC CCAGGCCCTC ACCCTGANGC TCCAGTGAC CATCAACAG
```

```
2601 TTCTTCGAGC CCACCGAGAT GCGGCGCCAG GATTTCCTCC AGCGCTGGAA
2651 GCAGCTGAGC CTCCTCAAC AGGAGCGCCA GAAATCTTC AAAGCCAACC
2701 ACCCATGGA CGCAGAAGTT ACTAAGGCCA AGCTTCTGGG GTTTGGCTCT
2751 GCTCTCTGG ACAATGTGA CCCCACCCCT GAGAAGCTCG TGGGGCGGGG
2801 GATCATCCAG ACTAAAGCCC TGCGGTGGG CTGTCTGCTT CGGCTGGAGC
2851 CCAATGCCCA GCGCCAGATG TACCGGTGA CCTTGCAC CAGCAAGGAG
2901 CCGCTCTCCC GTACCTGTG TGAGCTCTG GCACAGCAGT TCTGAGCCCT
2951 GGAATCTGCC CCGGGGATG TGGCGGCAC TGGCAGCCC CTGGACTGA
3001 GGCAGTTTGT GTGGATGGGG GACCTCCACT GGTGACAGAG AAGACACCAG
3051 GGTTTGGGGG ATGCTTGGGA CTTTCTCCG GCCTTTTGT TTTTATTTT
3101 TGTTCATCTG CTGCTGTTTA CATTCTGGGG GGTTAGGGGG AGTCCCTCTC
3151 CCTCCCTTTC CCCCCAAGC ACAGAGGGGA GAGGGGCCAG GGAAGTGGAT
3201 GTCTCCTCCC CTCCCACCC ACCTGTGTT AGCCCCCTCT ACCCTCTCCC
3251 CATCCAGGGG CTGTGTATTA TTGTGAGCGA ATAAACAGAG AGACGCTAAA
3301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AA
```

BLAST Results

No BLAST result

Medline entries

89155572:
Cloning of cDNAs encoding two related 100-kD coated vesicle proteins
(alpha-adaptins).

97431776:
Alpha-adaptin, a marker for endocytosis, is expressed in complex
patterns during Drosophila
development.

Peptide information for frame 3

ORF from 78 bp to 2942 bp; peptide length: 955
Category: strong similarity to known protein

```
1 MPAVSKGDGM RGLAVFISDI RNCKSKEAEI KRINKELANI RSKFKGDKAL
51 DGYSKKKYVC KLLFIFLLGH DIDFGHMEAV NLLSSNKYTE KQIGYLFISV
101 LVNSNSSELIR LINNAIKNDL ASRNPTFMCL ALHCIAVNGS REMGEAFAAD
151 IPRILVAGDS MDSVKQSAAL CLLRLYKASP DLVPMGEWTA RVVHLLNDQH
201 MGVVTAAVSL ITCLCKKNPD DFKTCVSLAV SRLSRIVSSA STOLQDYTTY
251 FVPAPMLSVK LLRLLOQYFP PEDAANKGRL VECLFVLRNK AQEPFASKRV
301 QMSNAKMAIL PETISLIHY DSEFWLLVRA CNQLGQFLON RETMLRYLAL
351 ESMCTLASSE FSHEAVKTHI DTVINALKTE RDVSVQRRAA DLYAMCORS
401 NAKQIVSEML RYLETADYAI REEIVLKVAI LAEKYAVDYS WYVDILNLI
451 RIAGDIVSEE VMYRVLQIVT NRDDVQGYAA KTVFEALQAP ACHENMVKVG
501 GYILGEFGNL IAGDPRSSPP VQFSLLSKF HLCSVATRAL LLSTYIKFIN
551 LFPETKATIQ GVLRAQSRLR NADVELQORA VEYLTLSVA STDVLATVLE
601 EMPFFPERES SILAKLRKKK GPGAGSALDD GRDPSSNDI NGMEPTPST
651 VSTPSPSADL LGLRAAPPPA APPASAGAGN LLVDVDFGPA AQPSLGPTPE
701 EAFSLSPGED IGPPIEADL LLMFKVCKNH GVLFENHQLQ IQVSEFRQN
751 LGRMYLFYGN KTSVQFQFNS PTVVHPGDLQ TQLAVQTKRV AAQVDGGAQV
801 QQVLNIECLR DFLTPPLLSV RFRYGGAPQA LTLKLPTVIN KFFQPTENAA
851 QDFFQWKQL SLPQQAQKI FKANHPMDAE VTRAKLLGFG SALLONVDPN
901 PENFVGAGII QTKALQVGCL LRLEPNAQAQ MYRLTLRTSK EPVSRHLCEL
951 LAQQF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1_20h13, frame 3

PIR:B30111 alpha-adaptin C - mouse, N = 1, Score = 3990, P = 0

PIR:S11276 alpha-adaptin c - rat, N = 1, Score = 3987, P = 0

SWISSPROT:ADAC_RAT ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2
ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA MEMBRANE
ADAPTOR HAZ/AP2 ADAPTIN ALPHA C SUBUNIT).., N = 1, Score = 3982, P = 0

SWISSPROT:ADAC_MOUSE ALPHA-ADAPTIN C (CLATHRIN ASSEMBLY PROTEIN COMPLEX
2 ALPHA-C LARGE CHAIN) (100 KD COATED VESICLE PROTEIN C) (PLASMA
MEMBRANE ADAPTOR HAZ/AP2 ADAPTIN ALPHA C SUBUNIT).. N = 1, Score =
3976, P = 0

TREMBL:AB020706_1 gene: "KIAA0899"; product: "KIAA0899 protein"; Homo
sapiens mRNA for KIAA0899 protein, partial cds.. N = 1, Score = 3932, P
= 0

>PIR:B30111 alpha-adaptin C - mouse
Length = 938

HSPs:

Score = 3990 (598.6 bits), Expect = 0.0e+00, P = 0.0e+00
Identities = 787/955 (82%), Positives = 858/955 (89%)

```
Query: 1 MPAVSKGDQMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60
      MPAVSKGDQMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
Sbjct: 1 MPAVSKGDQMRGLAVFISDIRNCKSKEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC 60

Query: 61 KLLFIFLLGHGHDIDFGHMEAVNLLSSNKYTEKIGYLFISVLVNSSELIRLINNAIKNDL 120
      KLLFIFLLGHGHDIDFGHMEAVNLLSSN+YTEKIGYLFISVLVNSSELIRLINNAIKNDL
Sbjct: 61 KLLFIFLLGHGHDIDFGHMEAVNLLSSNRYTEKIGYLFISVLVNSSELIRLINNAIKNDL 120

Query: 121 ASRNPTFMCLALHICIANVGSREMGAEAFADIPRILVAGDSMOSVKQSAALCLRLYKASP 180
      ASRNPTFM LALHICIANVGSREM EAFA +IP+ILVAGD+MDSVKQSAALCLRLY+ SP
Sbjct: 121 ASRNPTFMGLALHICIANVGSREMAEAFAGEIPKILVAGDTMDSVKQSAALCLRLYRTSP 180

Query: 181 DLVPMGEWTRVHVLLNDQHMVGVTAAVSLITCLCKKNPDFFKTCVSLAVSRLSRIVSSA 240
      DLVPMG+WT+RVVHLLNDQH+GVVTA SLIT L +KNP++FKT VSLAVSRLSRIV+SA
Sbjct: 181 DLVPMGDWTRVHVLLNDQHVGVTAAATSLITLAQKNPEEFKTSVSLAVSRLSRIVTSA 240

Query: 241 STDLDQDYTYFVPAPWLSVKLLRLLCYPPPEDAAVKGRLECELETVLNKAQEPKSKKV 300
      STDLDQDYTYFVPAPWLSVKLLRLLCYPPP D AV+GRL ECLET+LNKAQEPKSKKV
Sbjct: 241 STDLDQDYTYFVPAPWLSVKLLRLLCYPPP-DPAVGRLECELET+LNKAQEPKSKKV 299

Query: 301 QHSNARNAILFETISLIHYDSEPNLLVRACNOLGQFLOHRETNRLRYLALESMTCLASSE 360
      QHSNARNA+LFE TSLITH+DSEPNLLVRACNOLGQFLOHRETNRLRYLALESMTCLASSE
Sbjct: 300 QHSNARNAVLFEATSLIHHOEPNLLVRACNOLGQFLOHRETNRLRYLALESMTCLASSE 359

Query: 361 FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAQIVSEMLRYLETADYAI 420
      FSHEAVKTHI+TVINALKTERDVSVRQRA DLYAMCDRSNA+QIV+EML YLETADY+I
Sbjct: 360 FSHEAVKTHIETVINALKTERDVSVRQRAVDLLYAMCDRSNAQIVAEMLSYLETADYSI 419

Query: 421 REEIVLKVAILAEKYAVDYSWYVDTILNLIRIAGDYVSEEWYRVLQVTNRDDVQGYAA 480
      REEIVLKVAILAEKYAVDY+MYVDTILNLIRIAGDYVSEEWYRV+QIV NRDDVQGYAA
Sbjct: 420 REEIVLKVAILAEKYAVDYTWYVDTILNLIRIAGDYVSEEWYRVIQVTNRDDVQGYAA 479

Query: 481 KTVFEALQAPACHENMVVGGYILGEFGNLIAGDPRSSPPVQFSLHSHFHLCSVATRAL 540
      KTVFEALQAPACHEN+VKVGGYILGEFGNLIAGDPRSSP +QF+LHSHFHLCSV TRAL
Sbjct: 480 KTVFEALQAPACHENLVKVGYYILGEFGNLIAGDPRSSPLIQFNLHSHFHLCSVPTRAL 539

Query: 541 LLSTYIKFINLPETKATIQGLVLRAGSQLRNADVELQORAVEYLTLSVASTDVLATVLE 600
      LLSTYIKF+NLFP E KATIQ VLR+ SQL+NADVELQORAVEYL L S+VASTD+LATVLE
Sbjct: 540 LLSTYIKFVNLFPVETKATIQDVLRSDSQLKNADVELQORAVEYLRSTVASTDILATVLE 599

Query: 601 EMPFFPERESSILAKLKKFGPGAGSALDDGRDPSSNDINGMEPTP---STVSTPSPS 657
      EMPFFPERESSILAKLK+KKG P +L++ +R+ S D+NGG EP P S STPSPS
Sbjct: 600 EMPFFPERESSILAKLKKKGPSTVTOLETKRENSI-DVNGGPEFVPASTSAASTPSPS 658

Query: 658 ADLLGLAAPP-PAAPPASAGNLLVDVFDGPAQFSLGPTPEAFSLSPGEDIGPPIP 716
      ADLLGL A PP P PP S+G G LLVDVF A+ ++ P L+PG ED
Sbjct: 659 ADLLGLGAVPAPTGPSPSGGG-LLVDVFDSDAS--AVAP-----LAPGSEDN----- 704

Query: 717 EADELLNKFVCKNNGVLFENQLLQIGVSEFRONLGRMYLFYGNKTSVQFQNFSPTVVHP 776
      +FVCKNNGVLFENQLLQIG+KSEFRONLGRM++FYGNKTS QF NF+PT++
Sbjct: 705 -----FARFVCKNNGVLFENQLLQIGLSEFRONLGRMYFYGNKTSQFNLFTPTLCA 759

Query: 777 GDLOTQLAVQTKRVAQVDGGAQVQVNLNIECLRDFTPLPILLSVFRYGGAPQALTKLP 836
      DLQT L +QTK V VDGAQVQV+NIEC+ DF P+L+++FRYGG Q +++KLP
Sbjct: 760 DDLQTNLNLQTKPVDPTVDGGAQVQVNNIECISDFTEAPVLNIQFRYGGTFQNVSVKLP 819

Query: 837 VTINKFFQPTEMAQQDFQWRKQLSLPQQAQKIFRANHPMDAEVTKAKLLGFGSALLDN 896
      +T+KNFFQPTEMA+QDFQWRKQLS PQE Q IFKA HPMD E+TRAK++GFGSALL+
Sbjct: 820 ITLKNFFQPTEMASQDFQWRKQLSNPQEQVQNIKAKHPMDTEITKAKIIGFGSALLEE 879

Query: 897 VDPNPENFVGAGIIQTKALQVGCILLRLEPNAQAQMYRLTLRTSKEPVSRHLCELLAAQOF 955
      VDPNP NFGAGII TK Q+GCLLRLEPN QAQMYRLTLRTSK+ VS+ LCELL++QF
```

Sbjct: 880 VDPNPANFVGAGIHTKTTIGICLLRLLEPNLQAQMYRLTLRTSKDVTVSQRLCELLSEQF 938

Pedant information for DKFZphut1_20h13, frame 3

Report for DKFZphut1_20h13.3

[LENGTH] 955
[MW] 105361.97
[pI] 7.75
[HOMOL] PIR:A30111 alpha-adaptin A - mouse 0.0
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,
YBL037w] 5e-67
[FUNCAT] 08.19 cellular import [S. cerevisiae, YBL037w] 5e-67
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBL037w] 5e-67
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDR238c]
4e-04
[PIRKW] heterodimer 0.0
[PIRKW] transmembrane protein 1e-65
[PIRKW] membrane trafficking 0.0
[PIRKW] receptor 0.0
[SUPFAM] beta-adaptin 5e-16
[PROSITE] MYRISTYL 7
[PROSITE] IQ_MHC 1
[PROSITE] AMITATION 1
[PROSITE] CK2_PHOSPHO_SITE 11
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] PKC_PHOSPHO_SITE 15
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 6.81 %

SEQ MPAVSKGDMRGLAVFISDIRNCKSEAEIKRINKELANIRSKFKGDKALDGYSKKKYVC
SEG
PRD ccc

SEQ KLLFIFLLGHDIDFGHMEAVNLLSSNKYTEKQIGYLFISVLVMSNELIRLINNAIKNDL
SEG
PRD hhhhhhhcc

SEQ ASRNPTFMCLALHCIAVGSREMGFAADIPRLVAGDSMDSVKQSAALCLLRLYKASP
SEG
PRD ccc

SEQ DLVPHGEMTARVVHLLNDQHMGVVTAAVSLITCLCKKNPDDFKTCVSLAVSRLSRIVSSA
SEG
PRD ccc

SEQ STDLQDYTYFVPAPWLSVKLLRLQLQYPPPEDAAVKGRLVECLTVLNKAQEPKSKKV
SEG
PRD ccc

SEQ QHSNAKNAILFETISLTIHYDSEPNLLVRACNLGQFLQHRETNRLRYLALESMTLASSE
SEG
PRD ccc

SEQ FSHEAVKTHIDTVINALKTERDVSVRQRAADLLYAMCDRSNAKQIVSEMLRYLETADYAI
SEG
PRD ccc

SEQ REEIVLKVAIAEKYAVDYSWYVDILNLIRIAGDYVSEVWYRVLIQIVTNRDDVQGYAA
SEG
PRD hhh

SEQ KTVFEALQAPACHENMVVGGYILGEFGNLIAGDPRSSPPVQFSLHSHKFLHLSVATRAL
SEG
PRD hhh

SEQ LLSTYIKFINLFPETKATIQGVLRAGSQLRNADVELQQRAYEYLTLSVASTDVLATVLE
SEG
PRD hhh

SEQ EMPFFPERESSILAKLRKKGPGAGSALDDGRDPSSNDINGGHEPTPTSTVSTPSPSADL
SEG
PRD hcc

SEQ LGLRAAPPAPASAGAGNLLVDVFDGPAQPSLGTPPEAFSLSPGEDIGPPIPEADE
SEG xx

```
PRD      ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      LLNKEVCKNNGVLFEQQLIGVKSEFRQNLGRMYLFYGNKTSVQFQNFSPVTVHVGDLQ
SEG      .....
PRD      cccccccccccccchhhhhhhchhhhhccccccccccccccccccccccccccccccccchh

SEQ      TQLAVQTKRVAQVDGGAQVQVQVNLIECLRDFLTPPLLSVRFYGGAPQALTLLKLPVTIN
SEG      .....XXXXXXXXXXXXXXXX.....
PRD      hhhhhhhccccccccchhhhhhhcccccccccccccccccccccccccccccccccccccc

SEQ      KFFQPTDMAAQDFQRWKQLSLPQGEAQKIFKANHPMDAEVTKAKLLGFGSALLDNVDPH
SEG      .....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      PENFVGAGIIQTKALQVGCLLRLPNAQAQMYRLTLRTSKPEPSRHLCELLAQOF
SEG      .....
PRD      cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

Prosites for DKFZphute1_20h13.3

PS00001	760->764	ASN_GLYCOSYLATION	PDOC00001
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	163->166	PKC_PHOSPHO_SITE	PDOC00005
PS00005	189->192	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	297->300	PKC_PHOSPHO_SITE	PDOC00005
PS00005	379->382	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	470->473	PKC_PHOSPHO_SITE	PDOC00005
PS00005	787->790	PKC_PHOSPHO_SITE	PDOC00005
PS00005	819->822	PKC_PHOSPHO_SITE	PDOC00005
PS00005	832->835	PKC_PHOSPHO_SITE	PDOC00005
PS00005	935->938	PKC_PHOSPHO_SITE	PDOC00005
PS00005	938->941	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	368->372	CK2_PHOSPHO_SITE	PDOC00006
PS00006	379->383	CK2_PHOSPHO_SITE	PDOC00006
PS00006	470->474	CK2_PHOSPHO_SITE	PDOC00006
PS00006	482->486	CK2_PHOSPHO_SITE	PDOC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDOC00006
PS00006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS00006	636->640	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00006	938->942	CK2_PHOSPHO_SITE	PDOC00006
PS00007	388->395	TYR_PHOSPHO_SITE	PDOC00007
PS00007	411->419	TYR_PHOSPHO_SITE	PDOC00007
PS00007	434->443	TYR_PHOSPHO_SITE	PDOC00007
PS00008	202->208	MYRISTYL	PDOC00008
PS00008	508->514	MYRISTYL	PDOC00008
PS00008	561->567	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00008	826->832	MYRISTYL	PDOC00008
PS00008	908->914	MYRISTYL	PDOC00008
PS00009	630->634	AMIDATION	PDOC00009
PS00290	127->134	IG_MHC	PDOC00262

(No Pfam data available for DKFZphute1_20h13.3)

DKFZphutel_20ml1

group: cell cycle

DKFZphutel_20ml1 encodes a novel 225 amino acid protein with similarity to yeast sds22 and protein phosphatase-1 regulatory subunits.

sds22 is a regulatory polypeptide of protein phosphatase-1 that is required for the completion of mitosis in both fission and budding yeast. The novel protein seems to be a new regulator protein for protein phosphatase-1.

The new protein can find application in modulating/blocking the activity of protein phosphatase-1 and in modulating the cell cycle.

similarity to suppressor protein sds22

complete cDNA, complete cds, EST hits
localisation? only a part of the STS matches

Sequenced by AGOWA

Locus: /map="17"?

insert length: 5822 bp
Poly A stretch at pos. 5803, polyadenylation signal at pos. 5786

```
1 GGGCGCTTGG TTCCCAGCA ACCCGGAGAC GCGTCTGCTG CGTGAACCG
51 CCGAGTTCCC AGCGTTGAG AAGGAAATTT CTGGATCTGT TATCTGTGAG
101 GAGGCCACTC CGTTGACAGT TGTGTAAAC TCCTGCTGTT TCCCGAGCTC
151 CAACCTCTCT GCTTCTACAC AACACTATCA TCAGGAAATA CCGGGGGGAA
201 GATGAACCCG CCGTGCAACT CGATGGAGCC GAGGGTGATG GACGATGACA
251 TGCTCAAGCT GGCGTCTGGG GACGAGGGCC CCGAGGAGGA GCGCGGGCAG
301 CTGGCCAAAG AGGAGGGCAT CCTCTTCAAG GATGTCCTGT CCTGCAGCT
351 GGACTTTCCG AACATCCCTC GCATAGACAA CCTCTGGCAG TTTGAGAACT
401 TGAGGAAGCT GCAGCTGGAC AATAACATCA TTGAGAAGAT CGAGGGCCTG
451 GAGAACCTCG CACACTTGGT CTGGCTGGAT CTGTCTTTCA ACACATTGA
501 GACCATCGAG GCGCTGACA CACTGGTGA CCGGAGAGC CTGAGCTGT
551 TCAACAACCG GATCTCCAAG ATCGACTCCC TGAACGCCCT GGTCAAGCTG
601 CAGGTGTTGT CGCTGGGCAA CAACCGGATT GACAACATGA TGAACATCAT
651 CTACCTCCGG CGGTTCAAGT GCGTCCGGAC GCTCAGCCTC TTAGGAACC
701 CTATCTCTGA GGCAGAGGAT TACAAGATGT TCATCTGTGC CTACCTTCCT
751 GACCTCATGT ACCTGGACTA CCGGCGCATT GATGACCACA CAGCAAGTGT
801 CTCCTCTCTA GTCTCCCAAG CCTGTGAGAC AGATTCTCTA AGCCCCAGG
851 TTTCTTGGA AAGGGCATTT GAGAGTAGC TTCCCTGCC CACACTAGG
901 AGAGAAAGGG CAGTCCCTC TTCTTAATCC CTTTACTCTA CTCTGTACA
951 GTGATCCAG CAGCACCTTT GTAAGTACTG TTTTGTGTGC GTTCCAGGG
1001 GCCAGGCCCT TTCCACACAC TGTCCAGGG CCACCTCACA GCCATCCTGC
1051 ACTGTCTAGT TTCCAGATG AAGAAGCTGA GGAGGGCTGG GAGCAGTGGC
1101 TCACGCTGTG AATCCCAGCA CTTTGAGAGG CTGAGGCGGG AGGATCGCTT
1151 GAGCCAAGGA GTTCAAGACC AGCCTGGGCA ACATAGGGAG ACCCATCTC
1201 TACAGAAACT AGCAAAATTA GCCAGGTGTG CTGGACACACA CCACTAATCC
1251 TGCTTACTCA CAAGCCGAG GTAGAAAGAT CCGTTGAGC TAGAGTTTG
1301 AGCTGTCAGT GAATTAAGAA GATGCCATTG CACTCCAGCC TGGGCAACAG
1351 AGTGAAGAAA TTAAGAAATTT AAGAAAGAAA AGAAGTTGAG GAGGCCAAG
1401 GAGGCCAAGC AGCCAGGATC ACTGGCTCAA GGCCAAGCCA GGATTCACCC
1451 TAAGTTGGTG TCATCCAGG AGCAATATTA ACAGCTGAGC TCCAGAGGGA
1501 ACCAGGCCAT CAGAGGCTCA GGCCTGGCTC TCAGGGGCCAG AGTCAGGGCT
1551 GGAGGTAGAG ACCTGAGTGT CATCTGAGGA TTGCCAATTG GCAGTAGTTG
1601 AAGCCATGGT AAGGGTGGG TCACTGGGG CACATGGAGT GAGTGGGGG
1651 ACGGGGACTA AGTCTAGAG GTGCCAGCAT TCCTGGCCAG GTACAGGGGG
1701 ATGAGCCAGT GCGGTGGAGA GAGCCAAGGG CCAGACCCCTC GTGACCAACC
1751 CTATGGCCTC ACTCTACCTC TGTCTGTTG TCCTCCTTCC CTAAAGAGG
1801 GCCAGAAGGC CTGCTGAGGG GTGTTGGGAG TGAGAGAGCA AGTCTCTGT
1851 GGAGAACACC CAGTCTGGGG CGAGGGGAGC GCTCCATTGC TGTGGCTCCT
1901 GGCCTGGAGA TGCGCCCGGG AACCCAGGCC TGCCAGCCTG CTTCCGCTC
1951 CTCTCTGTCT TTCTCTGATT TCCTGCGCT CACAAAACC TGTGTAGGT
2001 CATCAGGAGA TGGGATTCT CATCCAGGAG ACCTCATGGC TTTACAGCC
2051 TTCATGCAGG CCCCCTGTGA ACACCCCTGC CCATGGCGGG GAGGCTGCAG
2101 CATGGCAGAG GCGGCATGGC AGAGGCGGTG TGGCTCGGAG GAACCTCTGG
2151 TAACAATGCC ACTCCGTTTC CCTGGTCAGA AAAAGCTTGC GGAGGCTAAG
2201 CACCAATACA GCATCGACGA GCTGAAGCAC CAGGAGAACC TGATGCAGGC
2251 CCAGCTGGAG GACGAGCAGG CGCAGCGGGA GGAGCTAGAG AAGCACAAGA
2301 CTGCGTTTGT GGAACACCTG AATGGCTCCT TCCTGTTTGA CAGCATGTAC
2351 CCTGAGGACT CAGAGGCAA CAATCTGCC TACCTGCTG GTCTGGTGA
2401 GCTCTCTGAG ACCTACAAGG ACAAGTTTGT CATCATCTGC GTGAATATT
2451 TTGAGTATGG CCTGAAACAG CAGGAGAAGC GGAAGACAGA GCTTGACACC
2501 TTCACTGAAT GTGTCCCTGA GGCCATCCAG GAAACACAGG AGCAGGGCAA
```



```

2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAAATGC
2651 AGTGCTGACA TCAGTGAGTT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAACTGTAGC CTTTCATGGGC
2801 ACGCCTCTGC TGGAAACGTC CAGCACGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCCTGT TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAAAGTAT TCTTTGGGAA AATGAGAGGT GAGCTCATTC TTCTGAAATG
2951 GTCCCCCTAT CCTGGAAGTC AGTGGGGAGA GGTTTTTGAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCACTT TGGGGACCAA
3051 ACAAAGACCC CTCCGCATTG CAGCCTGCAG TTGCCGCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAAATGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAGTCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCCAAGT GATAGCACCT AATCAATTCC TCACACGTCA AGTGACACCT
3301 CAGACATGGG ATAGATTTCC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAGCA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCAGC CTTTTCAGGC ACGGAGATTG CCTTTCAACA
3451 TCCAAACATT TCCAGAAACC CATGTGCCAT CCTACTGTGA TTAAGTGGTG
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCCT ATTTTATTTC
3551 ACGAGAACAG CACATACATG TGTGTTGAAA TTATGTGAGG TGCTCACTCT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCTGCCCC ACCTTGACAG
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCC
3701 TGGAGATCAC ATCTTCCATT GTTCTTCAA TCAACACTAA CTCCCATTG
3751 GGCCTTAGGT GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CCAGCCCGCT CCTACTGAGG GCCTCCTCTC TGTGAGGCAC
3851 CTTGCAAAGC ATTTTGTGTG AAGTGACTCA TTAACTCTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 TGGCTGTGTG GTGAAATGCC TTCCCAGAG TTACACAGAG TGCTGGAGCT
4001 GGGAAATCTG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TGTCACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACCAGG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTCACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGTCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAACTGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAAACA TGTGTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTAT CGAAAATGTC CAAAGCCTAT ATCCTTCTG TGATGACCTT
4501 CCGCATGGGG AGGTGCTACA GAGCCCTGCG GCTTGTCCCG GCCTCTGGAC
4551 AAAAGAATGT TCCACAGGGT CTGAGGAGGT TTCCCGACCC TCAGAACAAAT
4601 GATGGCCTGG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ATCTGTTTGA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTT
4701 ACTAGGAAAT GACGCCCTCT GTGTTGCAGG CAAGCACACT CTGGGGTTGA
4751 GGAACCCCCC ACGTGGAAGA CACTATAAGG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCTT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGGCAGGAG CCTCCATGCA GGAGGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGCCC
5101 AGGGTTCCGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TCGCCCTTGG CCCTTAAAGT CTGCTCTCCG CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGGG CTGCTCCTCC CGGGCCTTTG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCCTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCCTAAC GACCTGCGCG CGCTTTTGTG CGATAAAGAT ACGATTGTGA
5401 ATGCTGTCGG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCAG AATCAACTCT TGGTGTACAC GTTTAATAGA
5501 CAGGATTCAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCACATCC TAGACTAGAT GAATGTCAGC CACAGGAGCT TCTTCAAAAC
5651 ATAGCACCAG CCCCAGCCAG GAGAAGGAAG TGCACACGCC TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATTCT
5751 TCCCCACCC CTGGA AAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA

```

BLAST Results

Entry HS1292248 from database EMBL:
human STS SHGC-53917.
Score = 874, P = 3.3e-33, identities = 180/185

Medline entries

No Medline entry